University of California
Santa Barbara

How statistical and biological mechanisms shape the patterns and dynamics of aggregation in host-parasite systems

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Ecology, Evolution, and Marine Biology

by

Mark Quentin Wilber

Committee in charge:

Professor Cheryl J. Briggs, Chair
Professor Roger M. Nisbet
Professor William W. Murdoch

December 2017

The Dissertation of Mark Quentin Wilber is approved.

Professor Roger M. Nisbet

Professor William W. Murdoch

Professor Cheryl J. Briggs, Committee Chair

October 2017

How statistical and biological mechanisms shape the patterns and dynamics of aggregation in host-parasite systems

by

Mark Quentin Wilber

To my parents. For showing me how to appreciate the world around me.

Acknowledgements

I want to thank my advisor Cheryl Briggs for her many pieces of advise throughout my tenure as a graduate student, incredible feedback, and welcoming nature as a PhD advisor. I also want to thank Roger Nisbet and Bill Murdoch who helped channel my methodological infatuation into ecological questions and helped me find the connection between ecological theory and ecological realism my work. And finally, thanks to Dr. Gina Wimp, because without her encouragement and mentorship, I would never have started this process in the first place.

I also want to thank a number of students and collaborators who were critical in helping me develop many of the ideas presented in this dissertation, including John Harte, Pieter Johnson, Marm Kilpatrick, Justin Kitzes, Roland Knapp, Kate Langwig, Hamish McCallum, Erica Newman, Thomas Smith, Mary Toothman, Sara Weinstein, and Jacob Weverka.

Finally, to Erika. Thanks for keeping me sane and happy throughout this process.

Curriculum Vitæ

Mark Quentin Wilber

Education

2017	PhD in Ecology, Evolution, and Marine Biology (Expected), University of California, Santa Barbara
2017	M.A. Probability and Applied Statistics, University of California, Santa Barbara
2015	M.A., Ecology, Evolution, and Marine Biology, University of California, Santa Barbara
2011	B.S. Biology (Major), Mathematics (Minor), Georgetown University

Publications

Johnson, P. T. J. and **Wilber, M.**. 2017. Biological and statistical processes jointly drive population aggregation: using host-parasite interactions to understand Taylor's power law. *Proceedings of the Royal Society, B*, 2084: 20171388

Wilber, M., Knapp, R. A., Toothman, M., Briggs, C. J. Resistance, tolerance and environmental transmission dynamics determine host extinction risk in a load-dependent amphibian disease. 2017. *Ecology Letters*, 20: 1169-1181

Adams, A. J., Kuperberg, S. J., **Wilber, M.**, Grefsurd, M., Bobzien, S., Vredenburg, V. T., and Briggs, C. J. 2017. Drought, host density, sex, and bullfrogs influence fungal pathogen infection in a declining lotic amphibian. *Ecosphere*, 8: e01740

Wilber, M., Johnson, P. T. J., and Briggs, C. J. 2017. When can we infer mechanism from parasite aggregation? A constraint-based approach to disease ecology. *Ecology*, 98: 688-702

Wilber, M., Langwig, K., Kilpatrick, A. M., McCallum, H. I., and Briggs, C. J. 2016. Integral Projection Models for host-parasite systems with an application to amphibian chytrid fungus. *Methods in Ecology and Evolution*, 7: 1182-1194

Wilber, M., Weinstein, S., and Briggs, C. J. 2016. Detecting and quantifying parasite-induced host mortality from intensity data: Method comparisons and limitations. *International Journal for Parasitology*, 46: 59-66

Kitzes, J and **Wilber, M.**. 2015. `macroeco`: Reproducible ecological pattern analysis in Python. *Ecography*, 39: 361-367

Wilber, M., Kitzes, J., and Harte, J. 2015. Scale collapse and the emergence of the power law species-area relationship. *Global Ecology and Biogeography*, 24: 883-895

Langwig, K., Voyles, J., **Wilber, M.**, Bolker, B., Blehert, D., Briggs, C., Cheng, T., Collins, J., Fisher, M., Frick, W., Kilpatrick, M., Linder, D., McCallum, H., Murray, K., Puschedorf, R., Rosenblum, E. B., Toothman, M. 2015. Context-dependent conservation responses to wildlife disease. *Frontiers in Ecology and the Environment*, 13: 195-202

Voyles, J., Bolker, B., Blehert, D., Briggs, C., Cheng, T., Collins, J., Fisher, M., Frick, W., Kilpatrick, M., Langwig, K., Linder, D., McCallum, H., Murray, K., Puschedorf, R., Rosenblum, E. B., Toothman, M., and **Wilber, M.** 2014. Moving beyond too little too late: Managing emerging infectious diseases in wild populations requires international policy and partnerships. *EcoHealth*, 12: 404-407

Newman, E., Harte, M. E., Lowell, N., **Wilber, M.**, Harte, J. 2014. Empirical tests of within- and across-species energetics in a diverse plant community. *Ecology*, 95: 2815-2825.

Grants, awards, and fellowships

2015 – present	Mellichamp Fellowship
2015 – present	National Science Foundation Graduate Research Fellowship
2012 – 2016	University of California Regents Special Pre-Doctoral Fellowship
2016	Macroecology of Infectious Disease Student Grant
2014 – 2015	Worster Award for Graduate-Undergraduate Research
2013	Ecology, Evolutionary, and Marine Biology Departmental Fellowship
2012	Big East Institutional Scholar Athlete
2009 – 2011	Georgetown Howard Hughes Medical Institute Scholar
2011	Phi Beta Kappa Inductee
2011	Sigma Xi Inductee
2011	Sigma Xi Medal for Biology
2010	ESPN Academic All-American

Presentations and posters

2017	Managing disease-induced declines in Mountain-yellow legged frogs: insights from dynamical models, Mountain-yellow legged frog workshop at the San Diego Zoo, 11/08/2017
2017	When do resistance and tolerance dictate disease-induced extinction risk?, Ecology and Evolution of Infectious Disease, 6/26/2017
2017	The role of pathogen load, disease transmission, and host resistance and tolerance in disease-induced amphibian extinctions, University of Georgia, 1/16/2017
2017	Transmission, invasion and disease-induced extinction in amphibian populations, Annual Mellichamp Poster Session, University of California, Santa Barbara, 4/3/2017
2017	Revisiting the biological basis of Taylor's Power Law: using host-parasite interactions to reveal the drivers of aggregation, Ecological Society of America, 8/8/2017
2016	A constraint-based approach to parasite aggregation, Ecological Society of America, 8/8/2016
2016	Using host-parasite Integral Projection Models to explore temperature-dependent amphibian disease dynamics, University of California, Santa Barbara Graduate Student Symposium, 2/13/2016
2016	Integral Projection Models for host-parasite systems, Annual Mellichamp Poster Session, University of California, Santa Barbara, 4/8/2016
2015	Temperature-dependent dynamics of Bd-amphibian epizootics, Amphibian Pathogens Annual Meeting, 10/31/2015
2015	The emergence of the power law species-area relationship, Ecological Society of America, 8/14/2015
2015	A top-down approach for describing aggregation in host-parasite systems, Ecology and Evolution of Infectious Disease, 5/28/2017
2015	Amphibian disease and the structure of microbial communities, Annual Mellichamp Poster Session, University of California, Santa Barbara, 10/16/2015
2014	The effects of introduced trout on the species abundance distributions of Sierra Nevada aquatic communities, Ecological Society of America, 8/12/2014
2013	Macroeco: a macroecological analysis package for Python, Frontiers in Macroecological Theory, University of California, Berkeley, 1/25/2013
2011	The effects of habitat fragmentation on host-parasitoid interactions, Sigma Xi Seminar Series, Georgetown University, 4/25/2011

Abstract

How statistical and biological mechanisms shape the patterns and dynamics of aggregation in host-parasite systems

by

Mark Quentin Wilber

Few hosts have many parasites while many hosts have few parasites – this axiom of parasite ecology is known as parasite aggregation and is so pervasive that it is one of the few general laws in disease ecology. The propensity of parasites to be aggregated has important implications for both making inference about the mechanisms structuring a host-parasite interaction as well as predicting population-level host-parasite dynamics. In this dissertation I ask two questions: 1) How do the dynamics of parasite load and aggregation affect disease transmission, epidemics, and endemics in wildlife disease? 2) When can we make inference about the mechanisms structuring a host-parasite interaction from observed patterns of parasite load? I develop constraint-based theory for host-parasite systems to identify when patterns of parasite aggregation provide information about the mechanisms driving a host-parasite interaction. This approach shows that common patterns of parasite aggregation are highly constrained (i.e. predictable) by a simple set of constraints, providing a system-independent explanation for the ubiquitous pattern of parasite aggregation across host-parasite systems. However, despite the highly constrained nature of parasite aggregation, I show that particular mechanisms, such as parasite-induced host mortality, can lead to deviations from constraint-based theory and I develop statistical procedures to detect these deviations in cross-sectional parasite load data. While constraint-based theory focuses on static patterns of parasite aggregation,

parasite aggregation also influences host-parasite dynamics. I show that when parasite aggregation is consistent with the predictions from constraint-based theory, the ability of parasites to regulate the host population and stabilize the host-parasite equilibrium is significantly reduced, compared to the canonical assumption of fixed parasite aggregation. Finally, I develop a mathematical framework using Integral Projection Models (IPMs) to model parasite load dynamics when parasite load is a continuous variable. In combination with laboratory and mesocosm experiments, I apply this approach to an amphibian species infected with a fungal pathogen and show that disease-induced host extinction is far more sensitive to the load dynamics of the parasite than to the transmission dynamics in the system. This work highlights the importance of considering parasite load dynamics when developing strategies to mitigate disease-induced host declines. Broadly, this dissertation illustrates how both bottom-up, mechanistic approaches and top-down, statistical approaches can be used to provide unique insights into the mechanisms structuring consumer-resource interactions.

Contents

Introduction

The risk that a parasite poses to a host population depends on both the ability of a parasite to invade and the propensity of that parasite to cause disease-induced host declines. The within-host dynamics of a parasite are important for determining the outcome of both these events. While within-host infection dynamics are comprised of a multitude of interacting factors, such as host immunity (Woolhouse, 1992), intra-specific parasite interactions (Barbour and Pugliese, 2000), inter-specific parasite interactions (Fenton et al., 2010), and host behavior (Wilson et al., 2002), these interacting factors are often manifested in the dynamics of parasite abundance (i.e. load) within a host. In macroparasite infections, such as those from helminths and ectoparasitic arthropods (Anderson and May, 1979), the importance of parasite load on the population-level dynamics of hosts and parasites is well-known and it is typically accounted for in host-macroparasite models (e.g. Anderson and May, 1978, 1991; Dobson and Hudson, 1992; Morrill and Forbes, 2016). One of the reasons for the importance of parasite load in macroparasite systems is that host vital rates directly relating to host fitness (e.g. host survival rate and host reproductive rate) are often a function of parasite load. Therefore, it is important to consider not only whether a host is infected, but also the parasite load with which a host is infected. My dissertation seeks to answer two questions 1) How do the dynamics and patterns of parasite load affect disease transmission, epidemics, and

endemics in host populations? and 2) When can patterns of parasite load inform us about the dominant mechanisms driving host-parasite interactions?

The distribution of parasite loads across hosts in a population can have significant effects on host-parasite dynamics (Anderson and May, 1978; Rosà and Pugliese, 2002). For example, depending on the shape of the distribution of parasites across hosts, this distribution can promote or inhibit the ability of parasites to regulate the host population and the stability of the host-parasite equilibrium (Anderson and May, 1978; May and Anderson, 1978; Kretzschmar and Adler, 1993). Empirically, the distribution of parasites across hosts tends to be aggregated (Crofton, 1971; Shaw and Dobson, 1995). Colloquially, this means that few hosts have many parasites and many hosts have few parasites. Statistically, this means that the variance of parasite load across hosts in a population is often greater than the mean number of parasites per host and the distribution is highly right skewed. This pattern of parasite aggregation is so ubiquitous that it is one of the few general laws in disease ecology (Shaw and Dobson, 1995; Shaw et al., 1998; Wilson et al., 2002; Poulin, 2007, 2013).

In addition to affecting the dynamics of host-parasite systems, aggregated parasite distributions can also theoretically reflect something about the mechanisms driving a host-parasite interaction (Crofton, 1971; Adjei et al., 1986; Ferguson et al., 2011; Grear and Hudson, 2011). In Chapter 1, I develop a constraint-based theory for parasite ecology that identifies when host-parasite distributions may contain information about the mechanisms driving a host-parasite interaction and when they are simply a product of inevitable statistical constraints. This chapter shows that simple statistical constraints predict a large portion of the observed aggregation in host-parasite systems, such that attributing specific mechanism to empirical patterns of parasite aggregation is challenging. However, this work identifies that particular mechanisms, such as parasite-induced

host mortality, can leave a distinct signature on the host parasite distribution, even after accounting for the statistical constraints on the system. In Chapter 2, I develop a novel statistical method to detect parasite-induced host mortality from parasite distributions and discuss the benefits and limitations of this approach compared to previous approaches.

In Chapter 3, building on the results in Chapter 1, I ask the question: if parasite distributions are generally consistent with predictions from constraint-based theory, how does this affect the population-level dynamics of host-parasite systems? I extend the standard host-macroparasite model of Anderson and May (1978) to include parasite aggregation that follows the predictions of constraint-based theory and examine how this addition affects the ability of parasites to regulate host abundance (i.e. prevent hosts from increasing exponentially), suppress equilibrium host abundance, and stabilize the host-parasite equilibrium. I find that constraint-based aggregation significantly reduces the ability of parasites to regulate the host population and reduces the stability of the host-parasite equilibrium, compared to canonical host-parasite models that assume that parasite aggregation is fixed. This provides a potential theoretical rationale for the scarcity of empirical evidence that parasites are the dominant factor regulating and stabilizing the equilibria of host populations.

While aggregation and parasite load are key components of host-macroparasite models, models of microparasites often ignore parasite load (e.g. models of bacteria, viruses, and fungus where hosts often mount a strong immune response and parasite replication occurs directly within the host, Anderson and May, 1979). Instead, these models place hosts into discrete groups of, for example, (S)usceptible, (I)nfected, and (R)ecovered individuals. This is often a very reasonable simplification when within-host parasite dynamics are occurring quickly and there is not a strong, detectable relationship between

parasite load and host vital rates. However, some fungal parasites that are typically classified as microparasites show highly load-dependent dynamics (Briggs et al., 2010; Fisher et al., 2012; Langwig et al., 2017). For example, in some systems, amphibians infected with the chytrid fungus *Batrachochytrium dendrobatidis* (Bd) show highly load-dependent mortality, such that mortality rate increases rapidly when load increases above a certain threshold (Briggs et al., 2010; Stockwell et al., 2010; Vredenburg et al., 2010). However, unlike traditional macroparasites where parasite load is measured as a discrete count of parasites within a host, Bd load on an amphibian host is measured as a continuous variable via molecular analysis (Boyle et al., 2004). Therefore, to adequately model the population-level dynamics of hosts infected with parasites such as Bd, a new modeling framework is needed that accounts for continuous levels of parasite load.

In Chapter 4, I use Integral Projection Models (IPMs) to develop a framework for modeling host-parasite systems. Host-parasite IPMs link individual-based, continuous measures of parasite load to population-level dynamics. In this chapter, I show how the host-parasite IPM can be fit to standard data collected in amphibian-Bd systems and describe the population-level projections that can be made from the parameterized host-parasite IPM. In Chapter 5, I use the host-parasite IPM framework developed in Chapter 4 to understand the role of resistance, tolerance, and transmission in disease-induced host extinction. Using a mesocosm experiment on the amphibian host *Rana muscosa/sierrae* infected with Bd, I first estimate the transmission function for this amphibian-Bd system and show that it depends on both density-dependent host-to-host contacts as well as transmission from an environmental pool of Bd zoospores. Using this transmission function and an independent laboratory experiment, I then parameterize a temperature-dependent host-parasite IPM to understand the relative importance of transmission dynamics compared to host resistance and tolerance in driving disease-

induced extinction in this system. I find that while changes in transmission dynamics do affect the Bd-induced extinction risk of *R. muscosa* as predicted by previous theory (e.g. De Castro and Bolker, 2005; McCallum, 2012), equivalent changes in host resistance and tolerance have much larger effects on extinction. By explicitly modeling Bd load, I show that managing Bd load, rather than Bd transmission, can be more effective for mitigating disease-induced extinction risk in this amphibian species.

In conclusion, the five chapters in this dissertation contribute to our understanding of 1) how parasite aggregation can affect the dynamics of host-parasite interactions and 2) when patterns of parasite aggregation can inform us about the dominant mechanisms driving host-parasite systems. By combining theory, laboratory and mesocosm experiments, and field data, this work shows that the dynamics of parasite load can have significant and unexpected effects on disease-induced extinction, regulation, and stability dynamics in host-parasite systems. However, while these dynamical effects stem from underlying host-parasite interactions that in turn lead to aggregated host-parasite distributions, my work shows that the mechanisms themselves are often unidentifiable from observed patterns of parasite aggregation. Broadly, this work illustrates how both bottom-up, mechanistic approaches and top-down, statistical approaches can be used to provide unique insights into the processes structuring consumer-resource interactions.

References

Adjei, E. L., A. Barnes, and R. J. G. Lester. 1986. A method for estimating possible parasite-related host mortality, illustrated using data from *Callitetrarhynchus gracilis* (Cestoda: Trypanorhyncha) in lizardfish (*Saurida* spp.). Parasitology. **92**:227–243.

Anderson, R. M., and R. M. May. 1978. Regulation and stability of host-parasite

interactions: I. Regulatory processes. Journal of Animal Ecology **47**:219–247.

Anderson, R. M., and R. M. May. 1979. Population biology of infectious diseases: Part I. Nature **280**:361 – 367.

Anderson, R. M., and R. M. May. 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.

Barbour, A. D., and A. Pugliese. 2000. On the variance-to-mean ratio in models of parasite distributions. Advances in Applied Probability **32**:701–719.

Boyle, D. G., D. B. Boyle, V. Olsen, J. A. T. Morgan, and A. D. Hyatt. 2004. Rapid quantitative detection of chytridiomycosis (*Batrachochytrium dendrobatidis*) in amphibian samples using real-time Taqman PCR assay. Diseases of Aquatic Organisms **60**:141–8.

Briggs, C. J., R. A. Knapp, and V. T. Vredenburg. 2010. Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. Proceedings of the National Academy of Sciences of the United States of America **107**:9695–9700.

Crofton, H. D. 1971. A quantitative approach to parasitism. Parasitology **62**:179–193.

De Castro, F., and B. Bolker. 2005. Mechanisms of disease-induced extinction. Ecology Letters **8**:117–126.

Dobson, A. P., and P. J. Hudson. 1992. Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenuis* in red grouse. II. Population models. Journal of Animal Ecology **61**:487–498.

Fenton, A., M. E. Viney, and J. Lello. 2010. Detecting interspecific macroparasite interactions from ecological data: patterns and process. Ecology Letters **13**:606–15.

Ferguson, J. A., W. Koketsu, I. Ninomiya, P. A. Rossignol, K. C. Jacobson, and M. L. Kent. 2011. Mortality of coho salmon (*Oncorhynchus kisutch*) associated with burdens of multiple parasite species. International Journal for Parasitology **41**:1197–205.

Fisher, M. C., D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw, and S. J. Gurr. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature **484**:186–94.

Grear, D. A., and P. Hudson. 2011. The dynamics of macroparasite host-self-infection: a study of the patterns and processes of pinworm (Oxyuridae) aggregation. Parasitology **138**:619–27.

Kretzschmar, M., and F. R. Adler. 1993. Aggregated distributions in models for patchy populations. Theoretical Population Biology **43**:1–30.

Langwig, K. E., J. Hoyt, K. Parise, W. Frick, J. Foster, and A. M. Kilpatrick. 2017. Resistance in persisting bat populations after white-nose syndrome invasion. Philosophical Transactions of the Royal Society B **372**:20160044.

May, R. M., and R. M. Anderson. 1978. Regulation and stability of host-parasite population interactions: II. Destabilizing processes. Journal of Animal Ecology **47**:249–267.

McCallum, H. 2012. Disease and the dynamics of extinction. Philosophical Transactions of the Royal Society B **367**:2828–39.

Morrill, A., and M. R. Forbes. 2016. Aggregation of infective stages of parasites as an adaptation and Its implications for the study of parasite-host interactions. The American Naturalist **187**:225–235.

Poulin, R. 2007. Are there general laws in parasite ecology? Parasitology **134**:763–76.

Poulin, R. 2013. Explaining variability in parasite aggregation levels among host samples. Parasitology **140**:541–6.

Rosà, R., and A. Pugliese. 2002. Aggregation, stability, and oscillations in different models for host-macroparasite interactions. Theoretical Population Biology **61**:319–34.

Shaw, D. J., and A. P. Dobson. 1995. Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. Parasitology **111**:111–133.

Shaw, D. J., B. T. Grenfell, and A. P. Dobson. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology **117**:597–610.

Stockwell, M. P., J. Clulow, and M. J. Mahony. 2010. Host species determines whether infection load increases beyond disease-causing thresholds following exposure to the amphibian chytrid fungus. Animal Conservation **13**:62–71.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs. 2010. Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proceedings of the National Academy of Sciences **107**:9689–94.

Wilson, K., O. N. Bjoernstad, A. P. Dobson, S. Merler, G. Poglayen, A. F. Read, and A. Skorping, 2002. Heterogeneities in macroparasite infections: patterns and processes. Chapter 2, pages 6–44 *in* P. J. Hudson, A. Rizzoli, B. Grenfell, H. Heesterbeek, and A. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Woolhouse, M. E. 1992. A theoretical framework for the immunoepidemiology of helminth infection. Parasite Immunology **14**:563–78.

Chapter 1

When can we infer mechanism from parasite aggregation? A constraint-based approach to disease ecology

1.1 Abstract

Few hosts have many parasites while many hosts have few parasites - this axiom of macroparasite aggregation is so pervasive it is considered a general law in disease ecology, with important implications for the dynamics of host-parasite systems. Because of these dynamical implications, a significant amount of work has explored both the various mechanisms leading to parasite aggregation patterns and how to infer mechanism from these patterns. However, as many disease mechanisms can produce similar aggregation patterns, it is not clear whether aggregation itself provides any additional information about mechanism. Here we apply a "constraint-based" approach developed in macroecology that allows us to explore whether parasite aggregation contains any additional information beyond what is provided by mean parasite load. We tested two constraint-based null models, both of which were constrained on the total number of parasites P and hosts H found in a sample, using data from 842 observed amphibian host-trematode parasite distributions. We found that constraint-based models captured $\sim$85% of the observed variation in host-parasite distributions, suggesting that the constraints P and H contain much of the information about the shape of the host-parasite distribution. However, we also found that extending the constraint-based null models can identify the potential role of known aggregating mechanisms (such as host-heterogeneity) and disaggregating mechanisms (such as parasite-induced host mortality) in constraining host-parasite distributions. Thus, by providing robust null models, constraint-based approaches can help guide investigations aimed at detecting biological processes that directly affect parasite aggregation above and beyond those that indirectly affect aggregation through P and H.

1.2 Introduction

Disease ecology has traditionally emphasized mechanistic descriptions of infection patterns (Anderson and May, 1978; Duerr et al., 2003; Poulin, 2007). One particular pattern observed in macroparasites, such as parasitic helminths and arthropods that do not directly reproduce within their host (Anderson and May, 1979), is that many hosts in a population tend to have few parasites and a few hosts tend to have many. In statistical parlance this means that parasites tend to be aggregated within their hosts. This pattern is so ubiquitous in parasites that is has been called one of the few general laws in disease ecology (Poulin, 2007).

Canonical models of host-macroparasite dynamics have illustrated that a balance between parasite pathogenicity and parasite aggregation plays an important role in the ability of a parasite to regulate a host population (Anderson and May, 1978; Tompkins et al., 2002). In general, the stability of a host-parasite system and the regulation of a host population by parasites requires some level of parasite aggregation and that parasite pathogenicity is not too high (Anderson and May, 1978). Because of the importance of parasite aggregation, much empirical and theoretical work has sought to understand both the mechanisms that can lead to aggregation in host-macroparasite systems (Anderson and Gordon, 1982; Wilson et al., 2002; Raffel et al., 2011; Gourbière et al., 2015), and how to infer the dominant mechanisms structuring a host-parasite system from observed aggregation patterns (Duerr et al., 2003; Grear and Hudson, 2011; Wilber et al., 2016).

Traditionally, studies of macroparasite aggregation have relied on a process-based approach where various aggregating and disaggregating mechanisms are sequentially incorporated into unaggregated null models until observed levels of aggregation are obtained (Anderson and Gordon, 1982; Isham, 1995; Chan and Isham, 1998; Pugliese et al.,

1998; Rosà and Pugliese, 2002; Rosà et al., 2003; Grear and Hudson, 2011; Fowler and Hollingsworth, 2016). While the process-based approach has usefully illuminated various aggregating and disaggregating mechanisms in host-parasite systems (summarized in Wilson et al., 2002), it suffers from the "many-to-one problem" inherent in much of ecology (Frank, 2014): there are many process-based models that can result in similar levels of parasite aggregation making it difficult to identify the specific processes leading to aggregation from patterns alone. When lab or field experiments are not a viable option to identify mechanism in host-parasite systems, it would be useful to have some criteria to identify when observed patterns of parasite aggregation may provide some information about the mechanisms influencing a host-parasite system or when most of the information is provided in the mean parasite abundance.

Recently developed constraint-based models used in macroecology provide such a criteria. These models are different from the process-based approach in that they attempt to predict the most-likely form of a population- or community-level distribution using only a known set of statistical constraints (Harte, 2011; Locey and White, 2013; Newman et al., 2014; Xiao et al., 2015*b*). The constraint-based approach does not propose that biological mechanisms are not acting in a system; it contends that many different combinations of these mechanisms lead to similar patterns of aggregation with predictable statistical properties (Frank, 2009; McGill and Nekola, 2010; Frank, 2014). This is important because these models can then be used as robust null models (i.e. models that do not trivially fail) to identify when a given observed distribution contains biological information beyond that given by the constraints used to predict the distribution (Locey and White, 2013; Harte and Newman, 2014). Similarly, null model approaches have been used in community ecology to identify when signals of an ecological process can be discerned from observed changes in a community metric (e.g. changes in β diversity) when

a change in this metric can also be a concomitant result of changes in another metric (e.g. changes in α diversity; Chase and Myers, 2011; Chase et al., 2011). These constraint-based and null model approaches have had much success in understanding patterns in free-living populations and communities (Leibold and Mikkelson, 2002; White et al., 2012; Ulrich and Gotelli, 2013; Harte et al., 2015) and we argue that they can also be useful in addressing mechanistic questions about parasite aggregation in disease ecology.

For example, any observed host-parasite distribution is constrained by the total number of parasites P and the total number of hosts H in the sample. Given this, there are only a finite number of shapes that this sampled host-parasite distribution can take (i.e. the feasible set of the host-parasite distribution, Locey and White, 2013). If the shape of this observed host-parasite distribution is similar to the most likely distribution within this feasible set, then making inferences about the biological mechanisms leading to the shape of this distribution is difficult as the observed distribution is simply the most-likely distribution of all possible distributions (Haegeman and Loreau, 2009). In other words, many different combinations of host and/or parasite-related processes will lead to the same host-parasite distribution, and this is the distribution that is predicted by the constraint-based model.

This has important implications for disease ecology where one often wants to understand something about the mechanisms affecting a host-parasite system from the level of aggregation observed (Anderson and Gordon, 1982). Having some robust criteria for when a sampled host-parasite distribution shows "unusual" aggregation can help identify host-parasite systems where particular aggregating or disaggregating mechanisms are disproportionately constraining the distribution beyond the inherent (but biologically important) constraints imposed by P and H. We define "unusual" aggregation as a level of aggregation that is significantly different than the level of aggregation predicted

by the most likely distribution in the feasible set (e.g. Locey and White, 2013; Harte and Newman, 2014).

The traditional null model for the distribution of parasites across hosts follows a Poisson distribution, which is typically derived from a death-immigration process (Anderson and Gordon, 1982). The constraint-based approach for aggregation is asking a different question than tests of the traditional Poisson null model. Rejecting the Poisson null model indicates that this simple model is not capturing the important aggregating or disaggregating mechanisms in a host-parasite system, however failing to reject the Poisson null model is not proof that a system is following a simple death-immigration process. In contrast, the constraint-based null models make no assumptions about the particular processes leading to its predicted level of aggregation and simply predict the most likely level of aggregation for a system with P parasites and H hosts. Unlike the classic Poisson null hypothesis, failing to reject the constraint-based model tells us something important: our empirical pattern of aggregation does not contain much information about process beyond that already contained in P and H (Harte and Newman, 2014). Because this approach robustly identifies "unusual" aggregation, it can help us better understand when the effects of processes, such as parasite-induced mortality or host-heterogeneity, can be reliably inferred from observed host-parasite distributions.

This study has two goals. First, we use a dataset consisting of 22 unique amphibian host-trematode parasite pairings with over 8000 amphibians sampled at 205 sites over 5 years to test whether constraint-based models used in free-living systems also provide robust null models for host-parasite distributions. Second, we explore how, upon failing to describe host-parasite distributions, these constraint-based models can be extended to account for known aggregating mechanisms (such as host-heterogeneity) and disaggregating mechanisms (such as parasite-induced host mortality) in host-parasite systems.

We find that the shape of host-macroparasite distributions are generally well-predicted by the constraint-based approach. These results show that to reliably infer something about biological mechanism directly affecting patterns of parasite aggregation, one must first account for the strong constraints imposed on aggregation by P and H.

1.3 Methods

The methods section is organized as follows. The first section gives an overview of two constraint-based null models that have been recently used in the macroecological literature (Haegeman and Etienne, 2010; Locey and White, 2013; Xiao et al., 2015*a*). The second section describes how we generated predicted host-parasite distributions from these two constraint-based null models. The third section describes how we compared the constraint-based null models to data. Finally, the fourth section describes how we extended these constraint-based null models to account for known aggregating and disaggregating mechanisms in host-parasite systems. Table 1 contains a list of terms and definitions used to define the constraint-based models.

1.3.1 Defining the weighted feasible sets for constraint-based null models of parasite aggregation

The constraint-based null models that we consider have two constraints inherent in any sampled host-parasite distribution: the total number of parasites sampled P and the total number of hosts sampled H. Given these constraints, both models proceed by enumerating the feasible set of all possible macrostates of P parasites and H hosts (Locey and White, 2013). We define a macrostate as one possible unordered host-parasite distribution resulting from distributing P parasites among H hosts (Table 1).

For example, the feasible set of possible macrostates that we can observe given $P = 3$ parasites and $H = 3$ hosts is $F =\{\{3, 0, 0\}, \{2, 1, 0\}, \{1, 1, 1\}\}$. The macrostate {3, 0, 0} specifies that one host has three parasites and two hosts have zero parasites.

After specifying all of the possible macrostates in a feasible set constrained by P and H, each macrostate is then assigned a weight. For example, some macrostates may be combinatorially more likely to occur than others and thus will have a larger weight in the feasible set. Determining how to weight each macrostate depends on whether hosts and/or parasites are considered labeled or unlabeled (Table 1; Haegeman and Etienne, 2010).

One option is to specify that both hosts and parasites are unlabeled such that all possible macrostates are equally likely to occur. This is equivalent to integer partitions used in combinatorics (Bóna, 2006; Xiao et al., 2015*a*), so we call this model the "partition model". The partition model is process-independent and makes no assumptions about any potential mechanisms leading to a given macrostate (Locey and White, 2013). Therefore, no macrostate is more likely to occur than any other macrostate (Xiao et al., 2015*a*). Assuming that each macrostate is equally likely is not equivalent to assuming that any single host is equally likely to have a parasite abundance from zero to P. The probability of a single host having a parasite abundance of $x = 0, \ldots, P$ is $p(x|P, H) = \sum_{m \in F} p(x|m, P, H)p(m|P, H)$ where m is a macrostate in the feasible set F. Using our example from above, the partition model assigns each of the three macrostates in the feasible set an equal probability of 1 / 3. The probability of observing a single host with $x = 0, 1, 2,$ or 3 parasites is $p(0) = 3/9$, $p(1) = 4/9$, $p(2) = 1/9$, and $p(3) = 1/9$.

A second option for weighting macrostates is to again assume that parasites are unlabeled, but now assume that hosts are labeled. This is equivalent to integer compositions used in combinatorics (Bóna, 2006; Xiao et al., 2015*a*), so we call this model the

"composition model". Using the composition model, particular macrostates are more likely to occur because they are associated with a larger number of possible configurations. For example, given labeled hosts and unlabeled parasites, the macrostate {3, 0, 0} could be realized from three different configurations: (3, 0, 0), (0, 3, 0), and (0, 0, 3). Enumerating all the configurations for the other macrostates in our example feasible set, we see that the macrostate {2, 1, 0} can occur six ways and {1, 1, 1} can occur one way. Therefore, the macrostate {3, 0, 0} has a weight of 3 / 10, {2, 1, 0} has a weight of 6 / 10, and {1, 1, 1} has a weight of 1 / 10. The probability of observing a single host with $x = 0, 1, 2,$ or 3 parasites is $P(0) = 4/10$, $P(1) = 3/10$, $P(2) = 2/10$, and $P(3) = 1/10$.

More generally, for any macrostate m in a feasible set with P unlabeled parasites and H unlabeled hosts there are a total of $b_m = \frac{H!}{\Pi_{i \in A} h_i!}$ configurations with unlabeled parasites and labeled hosts (Brualdi, 2010). A is a set containing the unique parasite abundances found in macrostate m, i is a particular member of that set, and h_i is the number of hosts in macrostate m that have a parasite abundance i. Note that $\sum_{i \in A} h_i = H$. The total number of possible configurations of all macrostates using the composition model is given by $D = \frac{(H + P - 1)!}{P!(H - 1)!}$ (Harte, 2011). Taken together, the weight on any particular macrostate m using the composition model is $\frac{b_m}{D}$.

In summary, both the composition and partition models place our observed host-parasite distribution in the context of all possible observable host-parasite distributions. In particular, this allows us to ask an important question in parasite ecology: does a host-parasite distribution contain any information about biological mechanism beyond that already contained in P and H? Because there is no general consensus on which approach is preferable (Haegeman and Etienne, 2010; Xiao et al., 2015*a*), we consider both the partition model and the composition model in this study. Figure 1.1 gives a

visual comparison of these two models.

We could have also considered two other approaches: labeled hosts and labeled parasites or unlabeled hosts and labeled parasites. We chose not to consider these approaches because assuming labeled parasites is neither consistent with the pattern that we are interested in (i.e. the host-parasite distribution) nor how host-parasite systems are sampled. Assuming labeled parasites tracks the location of each individual parasite in the host population, whereas we are interested in the population-level distribution of parasites across hosts (see Xiao et al., 2015*a*, for the equivalent argument in free-living individuals). Moreover, specifying labeled parasites assumes that the system could be sampled by randomly choosing a parasite and assigning it a unique label and a label corresponding to the host in which it was found (assuming labeled hosts). This process would then be repeated until some number of P parasites were sampled. The total number of hosts H would then be given by the number of unique host labels on our P sampled parasites. This is not how host-parasite systems are sampled. Instead, H hosts are randomly sampled and the P parasites within these hosts are counted. This is more consistent with unlabeled parasites.

Despite these issues, the case with labeled hosts and labeled parasites is noteworthy because it results in a Poisson distribution of parasites across hosts (see Section 1.A). However, the model resulting in the Poisson distribution has exactly the same number of assumptions as the partition and composition models, so there is no *a priori* reason to favor one model over the other. The only way to discriminate between the approaches is to compare them to empirical host-parasite distributions (Haegeman and Loreau, 2009; Haegeman and Etienne, 2010; Xiao et al., 2015*a*), against which the Poisson almost universally fails (Shaw and Dobson, 1995; Shaw et al., 1998; Wilson et al., 2002). Fig. 1.S1 illustrates the completely unsurprising result that the Poisson distribution also does

not capture the level of parasite aggregation in the data we present here.

1.3.2 Moving from weighted feasible sets to constraint-based null model predictions

The proceeding section described how we enumerated and weighted the feasible sets for the partition and composition models. This section describes how we generate the predicted host-parasite distributions from these two models.

Given a weighted feasible set of macrostates from either the partition or composition model, the central tendency of this feasible set provides a prediction for the most likely host-parasite distribution given the constraints P and H (Locey and White, 2013). We define the central tendency of a weighted feasible set as the vector of marginal medians of this feasible set (Section 1.B). For most realistic values of P and H it is computationally intractable to enumerate all possible macrostates in the feasible set to compute this central tendency. To address this problem, we used the algorithms provided by Locey and McGlinn (2013) to randomly draw macrostates from all possible macrostates in a feasible set defined by P and H. We then computed the central tendency of this sample as an estimate of the central tendency of the full feasible set (Figure 1.1; Locey and White, 2013). To generate the predicted host-parasite distribution for the partition model, we drew 1000 random macrostates from a feasible set defined by P and H and used the central tendency of this sample as our predicted host-parasite distribution (Figure 1.1).

While we could have used the same approach to compute the predicted host-parasite distribution for the composition model by weighting each randomly drawn macrostate m by b_m/D, we instead used the analytical result from maximum entropy theory that the probability $p(x|P, H)$ of a single host having x parasites under the composition model

is (Haegeman and Etienne, 2010)

$$p(x|P,H) = \frac{\binom{P-x+H-2}{P-x}}{\binom{P+H-1}{P}} \tag{1.1}$$

The predicted rank abundance distribution of equation 1.1 is equivalent the central tendency of the weighted feasible set for the composition approach (Section 1.B). Moreover, equation 1.1 shows us that the composition approach is equivalent to assuming that host-parasite distributions follow a finite negative binomial distribution with $k = 1$ (Zillio and He, 2010). Note that we are not arbitrarily setting $k = 1$ – this is a direct result from maximizing entropy with respect to the constraints P unlabeled parasites and H labeled hosts. Moreover, the finite nature of this distribution is a direct result of the constraint P, which can lead to better descriptions of aggregation in finite populations (Zillio and He, 2010). However, P is similar to the number of trials in a binomial distribution and cannot be tuned to improved the fit of the model.

In summary, we used both a sampling based approach and an analytical formula to generate predicted host-parasite distributions from the weighted feasible sets of our two constraint-based null models. In the context of more commonly used distributions in disease ecology, these two constraint-based null models have one less parameter than a negative binomial model, which is a very flexible distribution that often fits host-parasite distributions very well (Shaw et al., 1998). We stress that the goal of this study is not to ask whether these distributions do better or worse than a negative binomial in predicting a host-parasite distribution, but whether host-parasite distributions tend to contain information beyond what is given by P and H.

1.3.3 Comparing constraint-based models to empirical data

Description of empirical data

To test whether empirical host-parasite distributions contained information beyond that given by the constraints P and H, we used an extensive dataset of all macroparasites found in 8099 amphibian hosts across 205 ponds (sites) in the East Bay region of California (Alameda, Contra Costa and Santa Clara counties) from 2009-2014 (Johnson et al., 2013). This included ponds from publicly accessible parks, open space preserves, municipal watershed districts, and private ranches. In this field study, we sampled recently metamorphosed amphibians, as these provide a reliable and standardized indicator of infections acquired during aquatic development from the associated pond. In a given survey event, we randomly collected at least 10 of each host species as they approached metamorphosis using the methods described in Johnson et al. (2016). To measure parasite abundance, we performed a systematic examination of all major tissues and organs in the sampled hosts for parasites (Hartson et al., 2011). The sampled amphibians consisted of *Pseudacris regilla* (Pacific chorus frog, $n = 4431$), *Anaxyrus boreas* (Western toad, $n = 1309$), *Lithobates catesbeianus* (American bullfrog, $n = 410$), *Taricha torosa* (California newt, $n = 1568$), and *Taricha granulosa* (Rough-skinned newt, $n = 381$).

We focused the following analyses on the five most common macroparasites in the system in terms of both prevalence and abundance. These were the larval trematodes *Ribeiroia ondatrae* (RION), *Echinostoma* sp. (ECSP), *Alaria* sp. (ALAR), *Cephalogonimus* sp. (CEPH), and *Manodistomum* sp. (MANO). All of these trematodes have complex life cycles in which their first intermediate hosts are pulmonate snails, their second intermediate host can be amphibians, snails or fishes, and their definitive hosts are water-associated vertebrates (reptiles, amphibians, birds, or mammals) (Johnson

and McKenzie, 2008).

Comparing models to data

We determined whether the empirical distributions of parasites across hosts deviated from the predictions of our two constraint-based models for each combination of host species and parasite species at each site during each year. We included a year-by-site-by-host-by-parasite distribution only if it had at least 10 parasites and 10 hosts. Given this criterion, we were able to compare the constraint-based models to 842 host-parasite distributions. As expected, 837 of these distributions were aggregated with a ln(variance to mean ratio) greater than zero (Fig. 1.S1). For each of these distributions, we extracted the total number of individuals of a given amphibian species (H) and parasites of a given trematode species (P) and calculated the corresponding rank abundance distribution (RAD) for the constraint-based models as the central tendency of the weighted feasible set (see *Moving from weighted feasible sets to constraint-based null model predictions*). The RAD gives the predicted parasite abundances from a given distribution for H hosts and assigns a rank of 1 to the host with highest abundance and a rank of H to the host with the lowest abundance (Harte, 2011; White et al., 2012).

To determine whether an observed host-parasite distribution deviated from the central tendency of a constraint-based model, we plotted the observed RAD (obs_i) versus the predicted RAD (pred_i). We then calculated the R^2 value based on a fit to the 1:1 line using the equation (White et al., 2012; Xiao et al., 2015*b*)

$$R^2 = 1 - \frac{\sum_i (\ln(\text{obs}_i + 1) - \ln(\text{pred}_i + 1)^2}{\sum_i (\ln(\text{obs}_i + 1) - \overline{\ln(\text{obs}_i + 1)})^2}$$

where i is the rank ($i = 1, \ldots, H$) of each observed or predicted host in a distribution. R^2 to the 1:1 line describes how much variance in the observed data is described by the

model prediction. If the model describes a large portion of the variation in the observed data then the R^2 value will be larger, with unity being a perfect prediction. If the model is a poor fit, the R^2 value will be much less then unity and possibly negative if the 1:1 line was a worse fit than assuming that each host had a parasite abundance equal to the mean of the observed distribution (White et al., 2012). We calculated R^2 values for each distribution independently as well as for all distributions combined. We also explored a number of alternative measures of goodness-of-fit that gave consistent results (Section 1.C).

1.3.4 Extending the constraint-based null models to account for aggregating and disaggregating mechanisms

When an observed host-parasite distribution deviated from the central tendency of a constraint-based null model, this provided evidence that additional constraints beyond just P and H were disproportionately affecting the system (Harte and Newman, 2014). We developed two ways to extend the constraint-based null models to detect whether classic aggregating and disaggregating mechanisms may be affecting host-parasite distributions beyond P and H.

Accounting for disaggregating mechanisms

Disaggregating mechanisms such as parasite-induced host mortality can play an important role in structuring empirically observed host-parasite distributions (Anderson and Gordon, 1982). The parasite *Ribeiroia ondatrae* is known to have a strong, intensity-dependent effect on the survival of some amphibian hosts where increased parasite intensity leads to increased limb-malformations and decreased survival (Johnson, 1999). This

means that hosts with large parasite burdens are removed from the system, making the parasite distribution more uniform. Therefore, *Ribeiroia*-induced host mortality may interact with P and H to further constrain the shape of host-*Ribeiroia* distributions.

We included *Ribeiroia*-induced host mortality as an additional constraint on the partition and composition null models. To do this, we used laboratory-derived survival curves that describe how *Ribeiroia* intensity affects amphibian host survival probability (Johnson et al., 2012). We focused on the amphibian species *Pseudacris regilla* because *Ribeiroia*-induced mortality and malformations in this species have been documented in the field and in the lab (Johnson, 1999; Johnson and McKenzie, 2008) and there were a large number of *P. regilla*-*Ribeiroia* distributions in the dataset on which to test the extended models ($n = 133$). The intensity-dependent survival curve specified the probability of an amphibian host surviving from larva to recent metamorph with some observed parasite intensity. We assumed that this curve followed a logistic function and estimated the parameters of this function from independent laboratory data (Figure 1.2A; Section 1.D; Johnson, 1999).

We then used this result to further constrain the partition and composition model predictions by assigning each macrostate a likelihood using the estimated survival function (Figure 1.2A). For each constraint-based model, the macrostates were then weighted by this likelihood such that macrostates with small likelihoods (e.g. ones that contained hosts with high parasite loads) were less likely to be observed than macrostates with large likelihoods. Using this weighting scheme, we sampled from models that were constrained on P, H and *Ribeiroia*-induced host mortality using a Metropolis-Hastings algorithm (Figure 1.2B; see Section 1.D for a full description of the algorithm used).

Once we obtained estimates of the mortality-constrained partition and composition models, we compared the resulting predictions to the observed *P. regilla*-*Ribeiroia* distri-

butions using the methods described in *Comparing models to data*. In addition, we also calculated an approximate AICc for the constraint-based model with and without an additional mortality constraint. We compared these models using ΔAICc where we considered an absolute value of $\Delta\text{AICc} > 2$ as evidence that one model was better than the other (Burnham and Anderson, 2002). The AICc values were approximate because there was no analytically defined likelihood for the mortality-constrained models. Therefore, we approximated the likelihood by drawing a large number of samples (e.g. 500 samples) from the mortality-constrained feasible set and computing the likelihood of a single host having x parasites using the equation $p(x|P,H) = \sum_{m \in \hat{F}} p(x|m,P,H)p(m|P,H)$, where $\hat{F}$ is the sampled feasible set. As we did not perform any additional model fitting to derive the mortality-constrained model, it was not statistically inevitable that the central tendencies of the mortality-constrained models would provide a better representation of the data. Therefore, an improvement in agreement between model and data, reflected in an increased R^2 or decreased AICc for the models with mortality relative to the models without mortality, is strong evidence that *Ribeiroia*-induced mortality is constraining the distribution beyond P and H.

Accounting for aggregating mechanisms

Host heterogeneity, whether it be in susceptibility, parasite encounter rates, behavior or other factors, is an important mechanism leading to aggregation in host-parasite systems (Cornell, 2010; Raffel et al., 2011). We accounted for this aggregating mechanism by extending the constraint-based models to include empirically observed levels of host heterogeneity. In particular, we explored discrete host heterogeneity where we assumed that overaggregation relative to the predicted model was a result of mixing discrete groups of hosts (Grafen and Woolhouse, 1993; Wilson et al., 2002). This approach is

different than the standard practice of fitting a negative binomial distribution to over-aggregated host-parasite distributions. If the goal of an analysis is to obtain the best possible fit to an observed host-parasite distribution then it is well-known that fitting a negative binomial distribution provides an excellent model of overaggregated host-parasite distributions (Shaw et al., 1998; Calabrese et al., 2011). However, if the goal of an analysis is to determine whether a host-parasite distribution contains any information beyond what is contained in P and H, fitting a negative binomial model does not provide immediate insight into what constitutes unusual aggregation or the potential host attributes leading to this overaggregation (but see Alonso and Pascual, 2006; Fowler and Hollingsworth, 2016, for various mechanistic interpretations of the negative binomial k parameter). Extending a constraint-based model to include discrete host-heterogeneity, as is done here, can help generate more specific hypotheses as to the relative importance of different levels of host-heterogeneity in structuring a host-parasite distribution.

To incorporate discrete host-heterogeneity, we used 5 observed host attributes by which we could bin hosts into groups of heterogeneity. The first attribute was host body size (i.e. snout-vent length), which is a well-known attribute affecting parasite exposure and aggregation (Grutter and Poulin, 1998; Poulin, 2013). The other 4 host attributes were the parasite abundances of the larval trematodes, excluding the focal trematode, infecting an individual host (see Fig. 1.3 for an example). Coinfection can potentially increase aggregation by increasing heterogeneity in host susceptibility to the focal parasite (Cattadori et al., 2008), but can also decrease aggregation by increasing intra-host parasite negative density dependence (Pacala and Dobson, 1988). Here we consider coinfection as a mechanism leading to increased aggregation.

Using these 5 host attributes, we used regression trees in which the response variable was the focal parasite abundance and the predictor variables were the 5 host attributes

described above (Fig. 1.3). Separate regression trees were run for each of the 842 host-parasite distributions. For a given host-parasite distribution, we found the best regression tree with 2-5 of groups of host heterogeneity and calculated the relative importance of each predictor variable based on how much they reduced the sum of squared error compared to the other predictor variables (Fig. 1.3). We restricted each group to have at least 2 hosts. Within each of these j groups we determined the total number parasites P_j and the total number of hosts H_j. The regression tree approach explores how various predictor variables affect mean parasite load (James et al., 2013), which is consistent with the constraint-based assumption that much of the information about the host-parasite distribution is contained in P and H.

To generate a constraint-based model RAD from the results of the regression tree, the RADs for each group j were computed with P_j and H_j and the predicted RAD was given by the concatenation for these j vectors (Fig. 1.3). This predicted mixture RAD could then be analyzed using the various methods described above. We also computed approximate AICc values for each heterogeneity model that we applied to an observed distribution. As described in the previous section, we did this by drawing 500 macrostates from the heterogeneity model to generate an estimate for the probability of a single host having x parasites under either the partition or composition assumptions. Finally, we employed a randomization test to ensure that any increase in R^2 after including host heterogeneity was due to the host attributes considered, rather than just the act of grouping itself (described in Section 1.E).

In summary, while P and H alone may sometimes not sufficiently constrain an observed host-parasite distribution, this approach is testing whether allowing P and H to vary as a function of host heterogeneity can account for deviations from the constraint-based null models. All analyses were performed in Python (version 2.7.11) and the code

to replicate the analysis can be found at `https://github.com/mqwilber/feasible_parasites`.

1.4 Results

1.4.1 Do host-parasite distributions contain information beyond that contained in P and H?

Overall, the partition model and composition model described 86% and 85% of the variation in all of the observed host-parasite distributions combined, respectively (Fig. 1.4A, B). For any particular host-parasite distribution, the median R^2 for the partition and composition models was 0.78 and 0.76, respectively (Fig. 1.4A, B). Examining the models with regard to host-by-parasite combinations, the median R^2 for the constraint-based models tended to be close to 80% for the various host-by-parasite combinations (Fig. 1.S2), with some notable exceptions for the host *Lithobates catesbeianus* and the parasites *Alaria* sp. and *Cephalogonimus* sp. (Fig. 1.S2-1.S5).

The partition model tended to describe more variation in host-parasite distributions than the composition model (Fig. 1.4, Fig. 1.S6). The composition model is equivalent to a finite negative binomial model with $k = 1$ and many of the observed host-parasite distributions in this study had maximum-likelihood estimated k parameters ($\hat{k}$) less than one (Fig. 1.S6). While $\hat{k} \neq 1$ is not necessarily incompatible with the composition model due to estimation error in the negative binomial k parameter (Lloyd-Smith, 2007), the partition model did predict more aggregated distributions than the composition model (Fig. 1.1). This led to the partition model accounting for a larger amount of the variance in host-parasite distributions with $k < 1$ (Fig. 1.S6).

1.4.2 Accounting for disaggregating and aggregating mechanisms

Disaggregating mechanisms: Parasite-induced host mortality

Including independently-estimated *Ribeiroia*-induced parasite mortality into the constraint-based models improved the overall fit of the models to *Pseudacris regilla-Ribeiroia* distributions. This was seen in three different metrics. First, there was a significant increase in the overall R^2 when the mortality constraint was included (bootstrapped 95% confidence interval for the difference in overall R^2 between the mortality constraint-based model and the null constraint-based model from 1000 re-samples: feasible set model, [0.019, 0.033]; maximum entropy model, [0.018, 0.035]; neither interval includes 0; Fig. 1.5A-D). This improvement in fit can be visualized by observing the tightening of the points to the 1:1 line when *Ribeiroia*-induced mortality was included in the model (Fig. 1.5A-D). Second, the median R^2 for individual *Pseudacris-Ribeiroia* distributions increased and the variance around the individual R^2 values decreased (Fig. 1.5B,D). Third, for individual distributions that were better fit under either the mortality or no mortality models based on the absolute value of ΔAICc > 2, a significant or marginally significant proportion were better under the mortality model (partition model: Binomial test, $N = 26$, better under mortality model $= 21$, $p = 0.002$; composition model: Binomial test, $N = 31$, better under mortality model $= 21$, $p = 0.07$; Fig. 1.5B,D).

Aggregating mechanisms: Host-heterogeneity

There were 124 unique host-parasite distributions that had $R^2 < 0.5$ and an observed variance to mean ratio greater than the variance to mean ratio of one of the constraint-based models. We considered these distributions to be overaggregated with respect to the constraint-based models. Of these 124 distributions, 48 were *Echinostoma* (12% of

all *Echinostoma* distributions) 29 were *Alaria* (35% of all *Alaria* distributions), 17 were *Cephalogonimus* (24% of all *Cephalogonimus* distributions), 17 were *Manodistomum* (33% of all *Manodistomum* distributions), and 13 were *Ribeiroia* (5% of all *Ribeiroia* distributions).

Considering only these overaggregated distributions, we used the regression tree analysis described above to test whether further constraining P and H based on known host attributes improved the fit of the constraint-based models to the observed host-parasite distributions. The heterogeneity models built from the regression tree analysis improved the fit of the constraint-based models to the empirical data beyond what would be expected by the inevitable increase in fit by simply grouping hosts (overall R^2 greater than the 95% interval from randomly permuting hosts into groups; Fig. 1.6). The improvement in fit can be visualized in Figure 1.6 by noting how the data points compress to the 1:1 line as more groups of heterogeneity are included. Note that this increase in model fit was not achieved by minimizing or maximizing any criteria about how well the heterogeneity model fit the observed host-parasite distribution. Finally, for three or more groups of host-heterogeneity, including just host body-size heterogeneity as an additional constraint yielded better models in terms of both higher R^2 values and larger AICc weights than including just heterogeneity in coinfection as an additional constraint (Fig. 1.7A-D).

1.5 Discussion

The shape of a sampled host-parasite distribution is necessarily constrained by the total number of hosts H and the total number of parasites P found in that sample (Haegeman and Etienne, 2010; Locey and White, 2013). While there are indisputably

biological mechanisms leading to the shape of this distribution (Wilson et al., 2002), inferring anything about these mechanisms may be difficult without first accounting for the constraints imposed by P and H. Here we use an extensive dataset of 22 host-parasite combinations and 842 empirical host-parasite distributions to show that aggregated host-parasite distributions tend to be consistent with the most likely distribution given P and H. This suggests that when trying to make inference about biological mechanism directly affecting patterns of parasite aggregation one must account for the mechanisms indirectly affecting aggregation through changes to P and H.

This finding has three important implications for disease ecology. First, there is a rich history in parasitology of using the shape of host-parasite distributions in combination with dynamic models and statistical techniques to infer which mechanisms may be affecting a given host-parasite system (Crofton, 1971; Anderson and Gordon, 1982; Grear and Hudson, 2011). While these approaches are in no way inappropriate, our results show that the instances in which the shape of a host-parasite distribution contains more information beyond what is contained in P and H may be more rare than previously thought. This result is consistent with other findings showing that log mean parasite load describes up to 88% of the variation in the log variance of parasite load, leaving only 13% of the variation to be described by biological mechanisms acting on something other than the mean (Shaw and Dobson, 1995; Poulin, 2013). Our results take this a step further by using constraint-based models to explicitly predict the entire host-parasite distribution given P and H. We find that, similar to Poulin (2013), much of the variability in the entire host-parasite distribution (not just the variance) is well predicted by mean parasite load and how many hosts are present in the sample. As a next step, explicitly considering whether specific attributes of observed host-parasite distributions systematically deviate from constraint-based predictions, such as the num-

ber of predicted uninfected hosts, could shed additional light on when host-parasite distributions contain much information beyond P and H.

The second implication is that the success of constraint-based models in disease ecology will allow them to be adopted as robust null models against which empirical host-parasite distributions can be compared. Constraint-based models are being increasingly used as robust null models in community ecology to determine when ecological mechanism may be disproportionately affecting the shape of population- and community-level distributions (Ulrich and Gotelli, 2013; Newman et al., 2014; Xiao et al., 2015*b*). In disease ecology, by using a constraint-based model to predict parasite aggregation given P and H, we can determine when a host-parasite system is showing unusual levels of aggregation to help direct modeling and experimental efforts. For example, future studies could explore whether factors such as the complexity of parasite life cycles (Lester and McVinish, 2016), self-reinfection processes (Grear and Hudson, 2011), or the composition of the host and parasite community in which a distribution is observed (Krasnov et al., 2006) lead to consistent deviations from constraint-based predictions.

Third, the general success of constraint-based models in describing host-parasite distributions has important implications for understanding the dynamics of host-macroparasite systems. Most macroparasite models explicitly model the state variables H and P (Anderson and May, 1978; Dobson and Hudson, 1992) and examine, in addition to other biological factors, how either fixed (Anderson and May, 1978) or dynamic aggregation (Kretzschmar and Alder, 1993; Rosà et al., 2003) influences host and parasite dynamics. Constraint-based models in turn predict that aggregation is largely determined by exactly these state variables. Therefore, a constraint-based approach to parasite ecology can be directly linked back to a more familiar mechanistic framework by examining the implications of constraint-based, aggregation predictions on dynamics of

the total number of hosts and parasites in a system. Linking constraint-based models for describing aggregation to dynamic equations for the state variables of a system has often been alluded to in macroecology (Supp et al., 2012; White et al., 2012), but has been difficult to implement (Harte, 2011). The rich empirical and theoretical understanding of biological factors affecting the total number of hosts and the total number of parasites in a system (Kretzschmar and Alder, 1993; Hudson et al., 1992; Dobson and Hudson, 1992) makes disease ecology an ideal field in which to make this connection.

In addition to providing robust null models and a unique opportunity to link dynamic, mechanistic models with a constraint-based approach, the constraint-based models can also be extended beyond null models to test the importance of potential aggregating and disaggregating mechanisms affecting host-parasite distributions. In this study, we extended the constraint-based models to include independently estimated relationships between parasite intensity and amphibian survival and found that accounting for the well-described negative effect of *Ribeiroia* on *P. regilla* (Johnson, 1999; Johnson et al., 2012) improved the fit of the constraint-based model to empirical host-parasite distributions. While this improvement in model fit was not drastic as P and H already accounted for 87% of the variation in the distributions, it was achieved using a survival curve estimated from an independent dataset (Johnson, 1999), providing strong evidence that parasite-induced mortality is influencing *P. regilla-Ribeiroia* distributions beyond just changes to P and H.

Moreover, we also found that extending constraint-based models to include heterogeneity in host body size and coinfection with other trematode parasites accounted for much of the overaggregation in observed distributions that were not well described by the constraint-based null models. In particular, we found evidence that host body size was generally a more important constraint on the host-parasite distribution than a host's

level of coinfection with other trematodes. This result is consistent with previous studies which have shown the importance of host age/body size heterogeneity for increasing parasite aggregation due to changes in host immunity and/or exposure to parasites with host age/body size (Pugliese et al., 1998; Poulin, 2013). Moreover, while previous work has shown that coinfection can act as a type of host heterogeneity and increase parasite aggregation (Cattadori et al., 2008), this same work has also shown that host characteristics such as age/body size, sex, and breeding status can often be more important factors affecting parasite aggregation and host-parasite dynamics than coinfection. While we have considered host-heterogeneity and parasite-induced mortality separately in this study, there is no reason that the constraint-based approach cannot be extended to included multiple mechanistic constraints. However, this must be done judiciously as imposing too many constraints could lead to trivial agreements between the model and data (Haegeman and Loreau, 2009).

In conclusion, constraint-based models provide a powerful framework for understanding when we can reliably infer mechanism from parasite aggregation. However, we are not advocating that the constraint-based approach should replace the process-based approach that has been so successful in disease ecology. Rather, the constraint-based approach is another tool in the disease ecologist's belt that can highlight when observed parasite aggregation is telling us something novel about the mechanisms acting in our system and when we should acknowledge the statistical inevitability that sometimes host-parasite distributions simply look how they must look given P and H.

1.6 Acknowledgments

We would like to thank the numerous members of the Johnson Lab at University of Colorado, Boulder who collected and processed the thousands of samples that comprise this dataset, as well as the many land managers who generously provided access to study sites, including East Bay Regional Parks, East Bay Municipal Utility District, Santa Clara County Parks, Hopland Research and Extension Center, Blue Oak Ranch Reserve, California State Parks, The Nature Conservancy, Open Space Authority and Mid-peninsula Open Space. We would also like to thank Justin Kitzes, Bill Murdoch, and Roger Nisbet for useful discussions and three anonymous reviewers for helpful comments on this manuscript. The National Institutes of Health (USA) Grant 1R01GM109499 from the Ecology of Infectious Disease program, the National Science Foundation (USA) (DEB-0841758, DEB-1149308), the National Geographic Society, and the David and Lucile Packard Foundation provided support for this work. M.W. was supported by a National Science Foundation, USA, Graduate Research Fellowship (Grant No. DGE 1144085) and the University of California Regents (USA).

References

Alonso, D., and M. Pascual. 2006. Comment on "A keystone mutualism drives pattern in a power function". Science **313**:1739; author reply 1739.

Anderson, R. M., and D. M. Gordon. 1982. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. Parasitology **85**:373–398.

Anderson, R. M., and R. M. May. 1978. Regulation and stability of host-parasite interactions: I. Regulatory processes. Journal of Animal Ecology **47**:219–247.

Anderson, R. M., and R. M. May. 1979. Population biology of infectious diseases: Part I. Nature **280**:361 – 367.

Bóna, M. 2006. A Walk Through Combinatorics: An Introduction to Enumeration and Graph Theory. Second edition. World Scientific Publishing Co., Toh Tuck LInk, Singapore.

Brualdi, R. 2010. Introductory Combinatorics. Fifth edition. Pearson Education, Inc, Upper Saddle River, New Jersey.

Burnham, K. P., and D. R. Anderson. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. Springer, New York.

Calabrese, J. M., J. L. Brunner, and R. S. Ostfeld. 2011. Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model. PloS one **6**:e29215.

Cattadori, I. M., B. Boag, and P. J. Hudson. 2008. Parasite co-infection and interaction as drivers of host heterogeneity. International Journal for Parasitology **38**:371–380.

Chan, M. S., and V. S. Isham. 1998. A stochastic model of schistosomiasis immuno-epidemiology. Mathematical Biosciences **151**:179–198.

Chase, J. M., N. J. B. Kraft, K. G. Smith, M. Vellend, and B. D. Inouye. 2011. Using null models to disentangle variation in community dissimilarity from variation in α-diversity. Ecosphere **2**:1–11.

Chase, J. M., and J. A. Myers. 2011. Disentangling the importance of ecological niches from stochastic processes across scales. Philosophical Transactions of the Royal Society B: Biological Sciences **366**:2351–2363.

Cornell, S. J., 2010. Modelling stochastic transmission processes in helminth infections. Chapter 5, pages 66–78 *in* E. Michael and R. C. Spear, editors. Modelling Parasite Transmission and Control. Springer-Verlag New York.

Crofton, H. D. 1971. A quantitative approach to parasitism. Parasitology **62**:179–193.

Dobson, A. P., and P. J. Hudson. 1992. Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenuis* in red grouse. II. Population models. Journal of Animal Ecology **61**:487–498.

Duerr, H. P., K. Dietz, and M. Eichner. 2003. On the interpretation of ageintensity profiles and dispersion patterns in parasitological surveys. Parasitology **126**:87–101.

Engmann, S., and D. Cousineau. 2011. Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. Journal of Applied Quantitative Methods **6**:1–17.

Fowler, A. C., and T. D. Hollingsworth. 2016. The dynamics of *Ascaris lumbricoides* infections. Bulletin of Mathematical Biology **78**:815–833.

Frank, S. A. 2009. The common patterns of nature. Journal of Evolutionary Biology **22**:1563–1585.

Frank, S. A. 2014. Generative models versus underlying symmetries to explain biological pattern. Journal of Evolutionary Biology **27**:1172–1178.

Gourbière, S., S. Morand, and D. Waxman. 2015. Fundamental factors determining the nature of parasite aggregation in hosts. Plos One **10**:1–17.

Grafen, A., and M. E. J. Woolhouse. 1993. Does the negative binomial distribution add up? Parasitology Today **9**:475–477.

Grear, D. A., and P. Hudson. 2011. The dynamics of macroparasite host-self-infection: a study of the patterns and processes of pinworm (Oxyuridae) aggregation. Parasitology **138**:619–27.

Grutter, A., and R. Poulin. 1998. Intraspecific and interspecific relationships between host size and the abundance of parasitic larval gnathiid isopods on coral reef fishes. Marine Ecology Progress Series **164**:263–271.

Haegeman, B., and R. S. Etienne. 2010. Entropy maximization and the spatial distribution of species. The American Naturalist **175**:E74–90.

Haegeman, B., and M. Loreau. 2009. Trivial and non-trivial applications of entropy maximization in ecology: A reply to Shipley. Oikos **118**:1270–1278.

Harte, J. 2011. Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics. Oxford University Press, Oxford, United Kingdom.

Harte, J., and E. A. Newman. 2014. Maximum information entropy: a foundation for ecological theory. Trends in Ecology & Evolution **29**:384–389.

Harte, J., A. Rominger, and W. Zhang. 2015. Integrating macroecological metrics and community taxonomic structure. Ecology Letters **18**:1068–1077.

Hartson, R. B., S. A. Orlofske, V. E. Melin, R. T. Dillon, and P. T. J. Johnson. 2011. Land use and wetland spatial position jointly determine amphibian parasite communities. EcoHealth **8**:485–500.

Hudson, P. J., D. Newborn, and A. P. Dobson. 1992. Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenuis* in red grouse. 1. Monitoring and parasite reduction experiments. Journal of Animal Ecology **61**:477–486.

Isham, V. 1995. Stochastic models of host-macroparasite interaction. The Annals of Applied Probability **5**:720–740.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. Introduction to Statistical Learning with Applications in R. Springer, New York, USA.

Johnson, P. T. 1999. The effect of trematode infection on amphibian limb development and survivorship. Science **284**:802–804.

Johnson, P. T. J., and V. J. McKenzie, 2008. Effects of Environmental Change on Helminth Infections in Amphibians: Exploring the Emergence of *Ribeiroia* and *Echinostoma* Infections in North America. Chapter 11, pages 249–280 *in* The Biology of Echinostomes. Springer-Verlag New York.

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

Johnson, P. T. J., J. R. Rohr, J. T. Hoverman, E. Kellermanns, J. Bowerman, and K. B. Lunde. 2012. Living fast and dying of infection: Host life history drives interspecific variation in infection and disease risk. Ecology Letters **15**:235–242.

Johnson, P. T. J., C. L. Wood, M. B. Joseph, D. L. Preston, S. E. Haas, and Y. P. Springer. 2016. Habitat heterogeneity drives the host-diversity-begets-parasite-diversity relationship: evidence from experimental and field studies. Ecology Letters **19**:752–761.

Krasnov, B. R., M. Stanko, D. Miklisova, and S. Morand. 2006. Host specificity, parasite community size and the relation between abundance and its variance. Evolutionary Ecology **20**:75–91.

Kretzschmar, M., and F. R. Alder. 1993. Aggregated distributions in models for patchy populations. Theoretical Population Biology **43**:1–30.

Leibold, M. A., and G. M. Mikkelson. 2002. Coherence, species turnover, and boundary clumping: elements of meta-community structure. Oikos **97**:237–250.

Lester, R. J. G., and R. McVinish. 2016. Does moving up a food chain increase aggregation in parasites? Journal of the Royal Society, Interface **13**:1 –11.

Lloyd-Smith, J. O. 2007. Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases. PLoS ONE **2**:1–8.

Locey, K. J., and D. J. McGlinn. 2013. Efficient algorithms for sampling feasible sets of macroecological patterns. PeerJ pages 1–23.

Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. Ecology Letters **16**:1177–85.

McGill, B. J., and J. C. Nekola. 2010. Mechanisms in macroecology: AWOL or purloined letter? Towards a pragmatic view of mechanism. Oikos **119**:591–603.

Newman, E. N., M. E. Harte, N. Lowell, M. Wilber, and J. Harte. 2014. Empirical tests of within- and across species energetics in a diverse plant community. Ecology **95**:2815–2825.

Niinimaa, A., and H. Oja, 2006. Multivariate Median. Pages 1 – 9 *in* Encyclopedia of Statistical Sciences. John Wiley & Sons, Ltd, Hoboken, New Jersey.

Pacala, S. W., and A. P. Dobson. 1988. The relation between the number of parasites/host and host age: population dynamic causes and maximum likelihood estimation. Parasitology **96**:197–210.

Poulin, R. 2007. Are there general laws in parasite ecology? Parasitology **134**:763–76.

Poulin, R. 2013. Explaining variability in parasite aggregation levels among host samples. Parasitology **140**:541–6.

Pugliese, A., R. Rosà, and M. L. Damaggio. 1998. Analysis of model for macroparasitic infection with variable aggregation and clumped infections. Journal of Mathematical Biology **36**:419–47.

Raffel, T. R., J. O. Lloyd-Smith, S. K. Sessions, P. J. Hudson, and J. R. Rohr. 2011. Does the early frog catch the worm? Disentangling potential drivers of a parasite ageintensity relationship in tadpoles. Oecologia **165**:1031–1042.

Rominger, A. J., and C. Merow. 2016. meteR : an r package for testing the maximum entropy theory of ecology. Methods in Ecology and Evolution pages 1–7.

Rosà, R., and A. Pugliese. 2002. Aggregation, stability, and oscillations in different models for host-macroparasite interactions. Theoretical Population Biology **61**:319–34.

Rosà, R., A. Pugliese, A. Villani, and A. Rizzoli. 2003. Individual-based vs. deterministic models for macroparasites: host cycles and extinction. Theoretical Population Biology **63**:295–307.

Scholz, F., and A. Zhu, 2015. kSamples: K-sample rank tests and their combinations.

Shaw, D. J., and A. P. Dobson. 1995. Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. Parasitology **111**:111–133.

Shaw, D. J., B. T. Grenfell, and A. P. Dobson. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology **117**:597–610.

Supp, S. R., X. Xiao, K. M. Ernest, and E. P. White. 2012. An experimental test of the response of macroecological patterns to altered species interactions. Ecology **93**:2505–2511.

Tompkins, D. M., A. P. Dobson, P. Arneberg, M. Begon, I. M. Cattadori, J. V. Greenman, J. A. P. Heesterbeek, P. J. Hudson, D. Newborn, A. Pugliese, A. P. Rizzoli, R. Rosa, F. Rosso, and K. Wilson, 2002. Parasites and host population dynamics. Chapter 3, pages 45–62 *in* P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heessterbeck, and A. P. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Ulrich, W., and N. J. Gotelli. 2013. Pattern detection in null model analysis. Oikos **122**:2–18.

White, E. P., K. M. Thibault, and X. Xiao. 2012. Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model. Ecology **93**:1772–8.

Wilber, M. Q., S. B. Weinstein, and C. J. Briggs. 2016. Detecting and quantifying parasite-induced host mortality from intensity data: Method comparisons and limitations. International Journal for Parasitology **46**:59–66.

Wilson, K., O. N. Bjoernstad, A. P. Dobson, S. Merler, G. Poglayen, A. F. Read, and A. Skorping, 2002. Heterogeneities in macroparasite infections: patterns and processes. Chapter 2, pages 6–44 *in* P. J. Hudson, A. Rizzoli, B. Grenfell, H. Heesterbeek, and A. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Xiao, X., K. J. Locey, and E. P. White. 2015*a*. A process-independent explanation for the general form of taylor's law. The American Naturalist **186**:E51–E60.

Xiao, X., D. J. McGlinn, and E. P. White. 2015*b*. A strong test of the Maximum Entropy Theory of Ecology. The American Naturalist **185**:E70–80.

Zillio, T., and F. He. 2010. Modeling spatial aggregation of finite populations. Ecology **91**:3698–3706.

Table 1.1: Definitions of terms used to describe the constraint-based null models.

Term	**Definition**
Labeled	Hosts or parasites are distinguishable
Unlabeled	Hosts or parasites are indistinguishable
Macrostate	Unordered vector of unlabeled parasite abundances. e.g. Given $P = 3$ and $H = 2$, the vector $\{3, 0\}$ is a macrostate
Configuration	Ordered vector of unlabeled parasite abundances. e.g. Given $P = 3$ an $H = 2$, the macrostate $\{3, 0\}$ has two configurations: (3, 0) and (0, 3)
Feasible set	All possible macrostates given P and H. e.g. The feasible set given $P = 3$ and $H = 2$ is $\{\{3, 0\}, \{2, 1\}\}$
Weighted feasible set	Feasible set in which macrostates have particular weights.
Partition model	Weights each macrostate in the feasible set by assuming unlabeled hosts and parasites. All macrostates have equal weights.
Composition model	Weights each macrostate in the feasible set by assuming labeled hosts and unlabeled parasites. Analogously, each macrostate can be realized by multiple configurations.

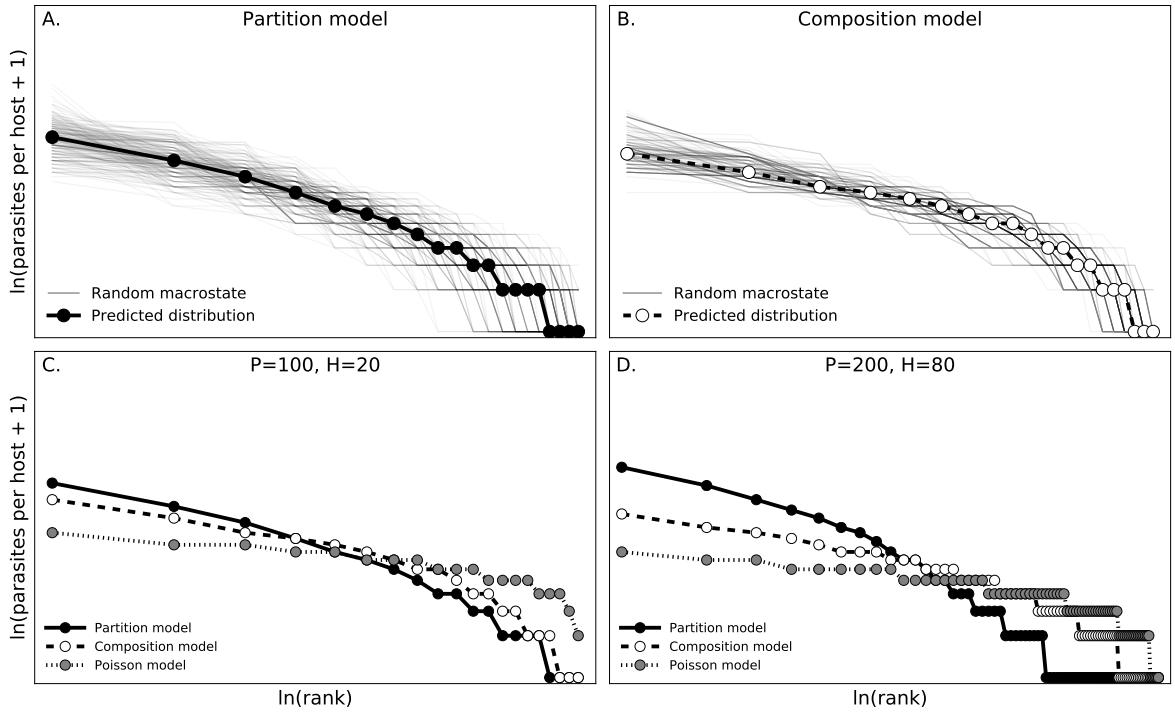


Figure 1.1: **A.** and **B.** show how the predictions from the partition model and composition model were generated. Random macrostates were drawn from the weighted feasible set (light gray lines) and the predicted distribution was computed as the central tendency of these randomly drawn weighted macrostates (thick lines with dots). Each dot represents a host in the predicted distribution with a given parasite abundance and rank. Hosts with low ranks (low ln(ranks)) have a larger number of parasites than hosts with high ranks (high ln (ranks)). **C.** and **D.** compare the partition and composition models for two different values of P and H. The more familiar Poisson model is also included for reference. The partition and composition models predict more aggregated host-parasite distributions than the Poisson model and the partition approach tends to produce more aggregated distributions than the composition model. The degree that the predictions from the partition and composition models differ depends on the values of P and H.

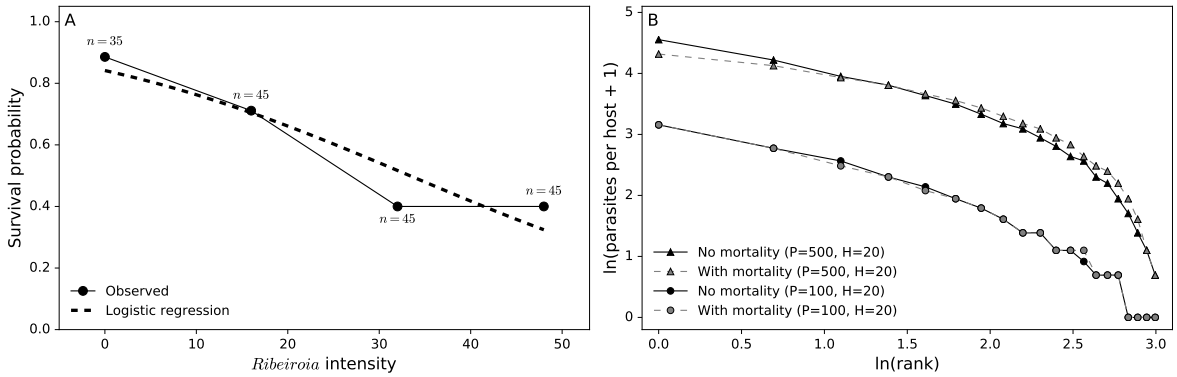


Figure 1.2: **A.** The black line and dots give the parasite-induced host mortality data from Johnson (1999) where *Pseudacris regilla* hosts were infected with varying *Ribeiroia* intensities. Frogs were exposed to *Ribeiroia* cercariae as tadpoles and the experiment was stopped after tadpoles metamorphosed. The dashed line gives the mean predictions of the logistic regression model fit to the data. This logistic regression was then imposed as a mortality constraint on the partition model and the composition model. **B.** An example of the effect of including the laboratory-estimated survival curve on the predictions of partition models with $P = 500$ parasites and $H = 20$ hosts and $P = 100$ and $H = 20$. The symbols indicate a given host in a predicted distribution with a particular parasite abundance and rank. Hosts with low ranks (low ln(ranks)) have a larger number of parasites than hosts with high ranks (high ln (ranks)). Depending on the values of P and H, the mortality constraint could noticeably reduce aggregation (triangles) or have little effect on aggregation (circles).

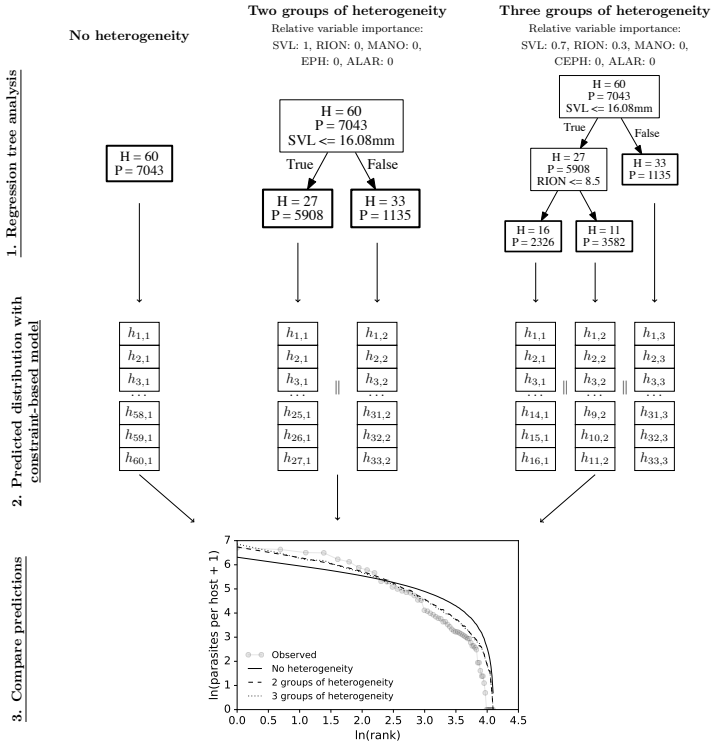


Figure 1.3: A diagram showing how host-heterogeneity can be incorporated into constraint-based models. **Step 1:** Consider, for example, a distribution for the parasite *Echinostoma* sp. in the host *Pseudacris regilla* with $H = 60$ hosts and $P = 7043$ parasites. When no host heterogeneity is included, the central tendency of the constraint-based model can be computed directly from H and P as described in the main text. To include groups of heterogeneity, a regression tree analysis is performed in which the response variable is *Echinostoma* abundance and the predictor variables are *P. regilla* body size (snout-vent length, SVL) and the abundance of *Ribeiroia ondatrae* (RION), *Alaria* sp. (ALAR), *Cephalogonimus* sp. (CEPH), and *Manodistomum* sp. (MANO) in a particular host. In the example above, the regression tree analysis shows that the "best" way to make two groups of heterogeneity given the predictor variables is to split the 60 *P. regilla* individuals into those with SVL $\leq$ 16.08 mm and those with SVL $>$ 16.08 mm. To make three groups of heterogeneity, *P. regilla* individuals with SVL $\leq$ 16.08 are again split into individuals with RION abundance $\leq$ 8.5. For each of these regression trees, we can determine the relative importance of each variable in building the regression tree by how much they decrease the sum of squared error compared to the other predictors. **Step 2:** We can then compute the central tendency of the constraint-based model for each of these groups of heterogeneity (the bold boxes above) using the total number of hosts and parasites in each heterogeneity group. Each heterogeneity group has its own rank abundance distribution with $h_{i,j}$ being the ith ranked host with some number of parasites in heterogeneity group j. Concatenating ($\|$) these rank abundance distributions together and re-ordering the resulting vector gives the predicted constraint-based model after allowing for P and H to vary with host heterogeneity. **Step 3:** These predicted distributions can then be compared to the observed host-parasite distribution.

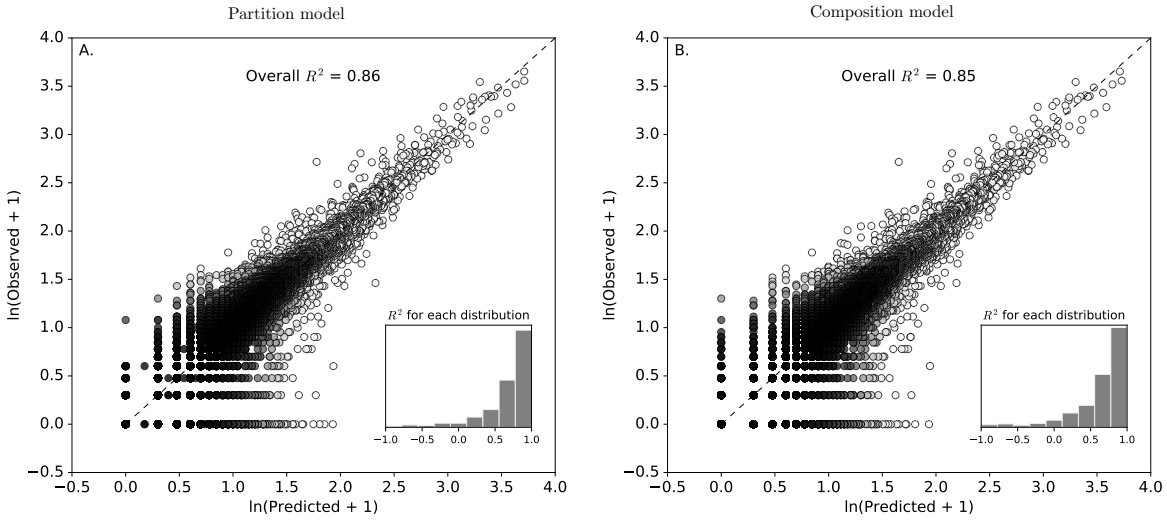


Figure 1.4: Plots showing the fit of the partition model **A.** and the composition model **B.** to all of the 842 observed host-parasite distributions considered in this study. The black-dashed line gives the 1:1 line and the overall R^2 gives the percent of variation that the constraint-based models explain in all of the observed data. Each point represents a single individual host from one of the 842 distributions with a given predicted and observed parasite abundance. Darker colors indicate a higher density of points in the region than lighter colors. The inset histogram shows the distribution of R^2 values calculated for each of the 842 individual host-parasite distributions.

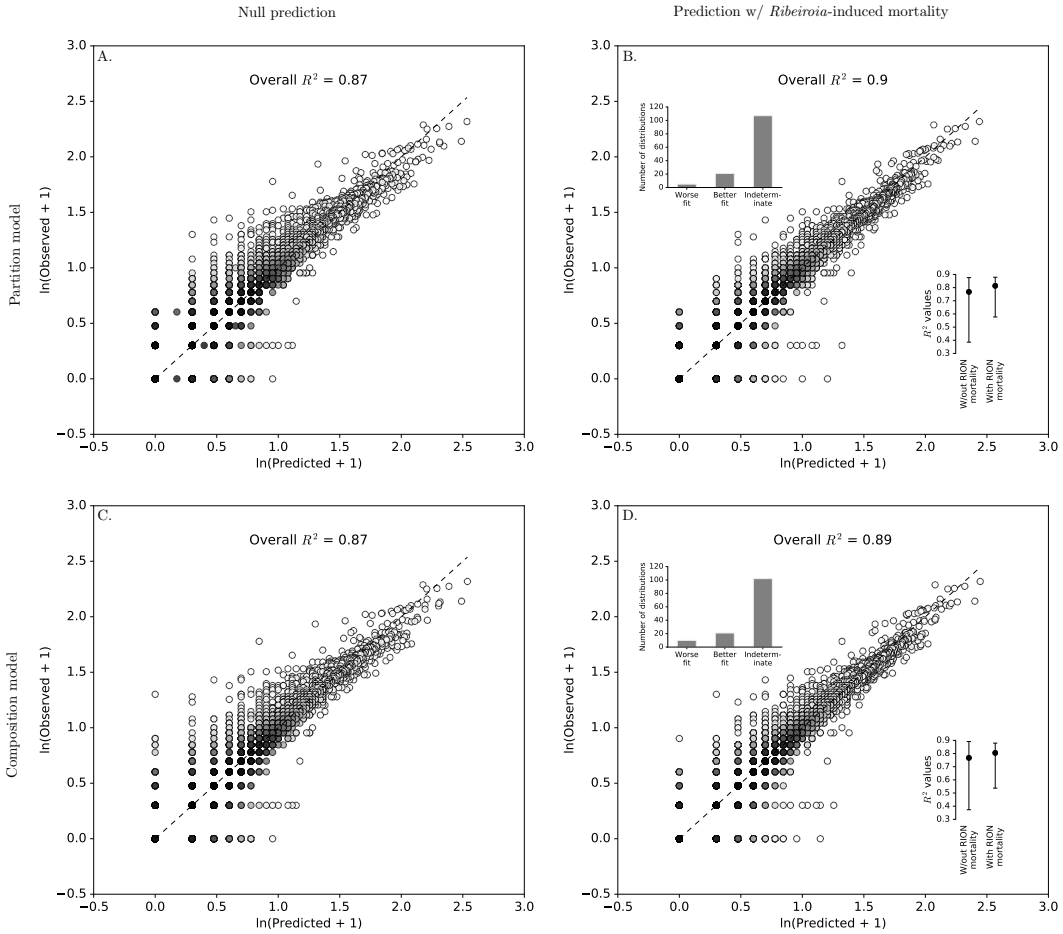


Figure 1.5: The effect of including empirically-estimated *Ribeiroia*-induced *Pseudacris regilla* mortality into the partition and composition models. The first column in this plot (**A.**, **C.**) compares 133 observed rank abundance distributions (RAD) of *Ribeiroia*-*P. regilla* with the RADs predicted by the constraint-based models before they were constrained on parasite-induced host mortality. The second column (**B.**, **D.**) compares the observed and predicted RADs after they were constrained on parasite-induced host mortality. The 1:1 line is given by the black, dashed line. Each point represents a single host with a given predicted and observed parasite abundance. Darker colors indicate a higher density of points in the region than lighter colors. The inset dot plots in B. and D. show the first through third quartiles of the R^2 values from each of the 133 distributions without and with *Ribeiroia*-induced mortality. The inset bar plots in B. and D. give the number of observed distributions that were a worse fit with the mortality constraint, a better fit with the mortality constraint, or were equally good with either (Indeterminate) based on ΔAICc.

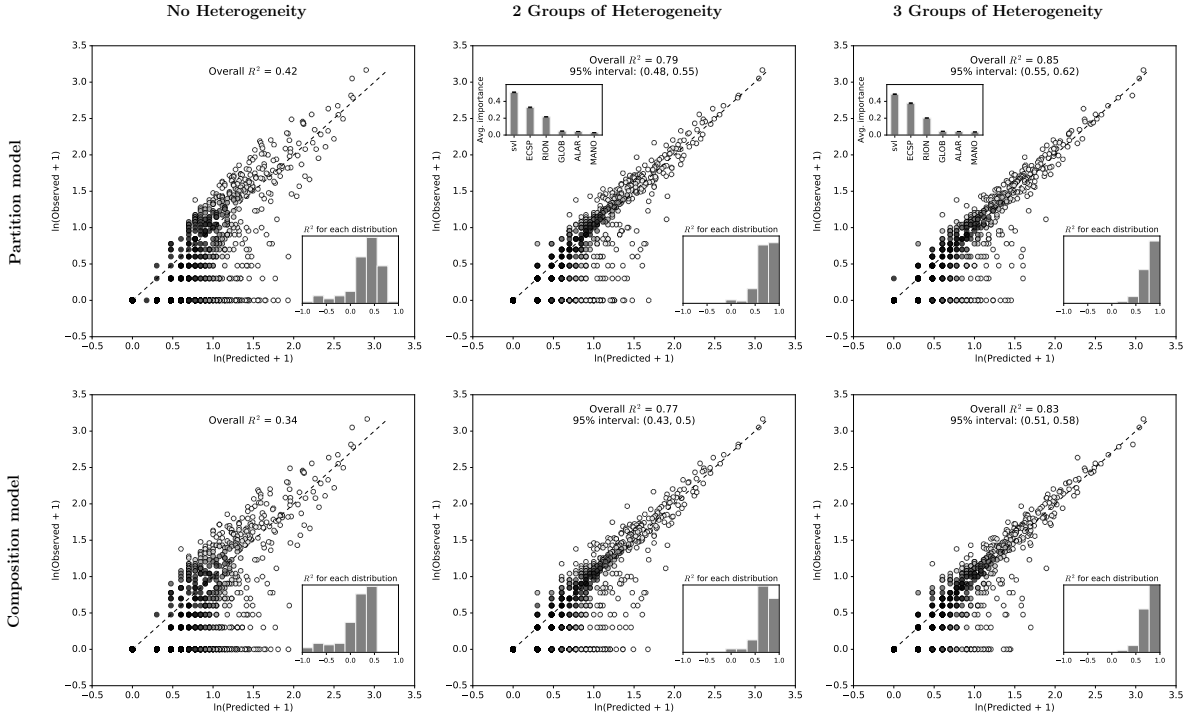


Figure 1.6: The effect of discrete heterogeneity on the 124 host-parasite distributions that were overaggregated relative to at least one of the constraint-based models (all hosts and parasites shown together). The first column in this plot shows the predicted rank abundance distributions (RAD) compared to the observed RADs when no host heterogeneity was included in either of the two constraint-based models. The black, dashed line gives the 1:1 line and the overall R^2 describes the amount of variation the constraint-based models described in all 124 overaggregated distributions. Each point represents a single host with a given predicted and observed parasite abundance. Darker colors indicate a higher density of points in the region than lighter colors. The histogram in the lower right hand side gives the distribution of R^2 values for each particular host-parasite distribution. The second and third columns in this plot show the effect of adding 2 and 3 groups of host heterogeneity, respectively, on the predicted host-parasite distributions based on the results from a regression tree analysis on known host attributes in the dataset. The plots in the upper left hand corner show the mean importance of a given host attribute in structuring the regression tree for all the 124 overaggregated host-parasite distributions. The predictor variables were body-size (svl), *Echinostoma* sp. (ECSP), *Ribeiroia* (RION), *Cephalogonimus* (CEPH), *Alaria* (ALAR), and *Manodistomum* (MANO). The predictor importance was the same for all models within a heterogeneity group and are therefore only displayed once for each group. Finally, the 95% interval displayed in the plot gives the 95% quantiles of overall R^2 values based on randomly permuting parasites into the groups predicted by the regression tree analysis. If the overall R^2 is greater than the interval, it shows that the increase in R^2 from the regression tree is a result of the predictors used in the regression tree analysis, rather than just grouping itself.

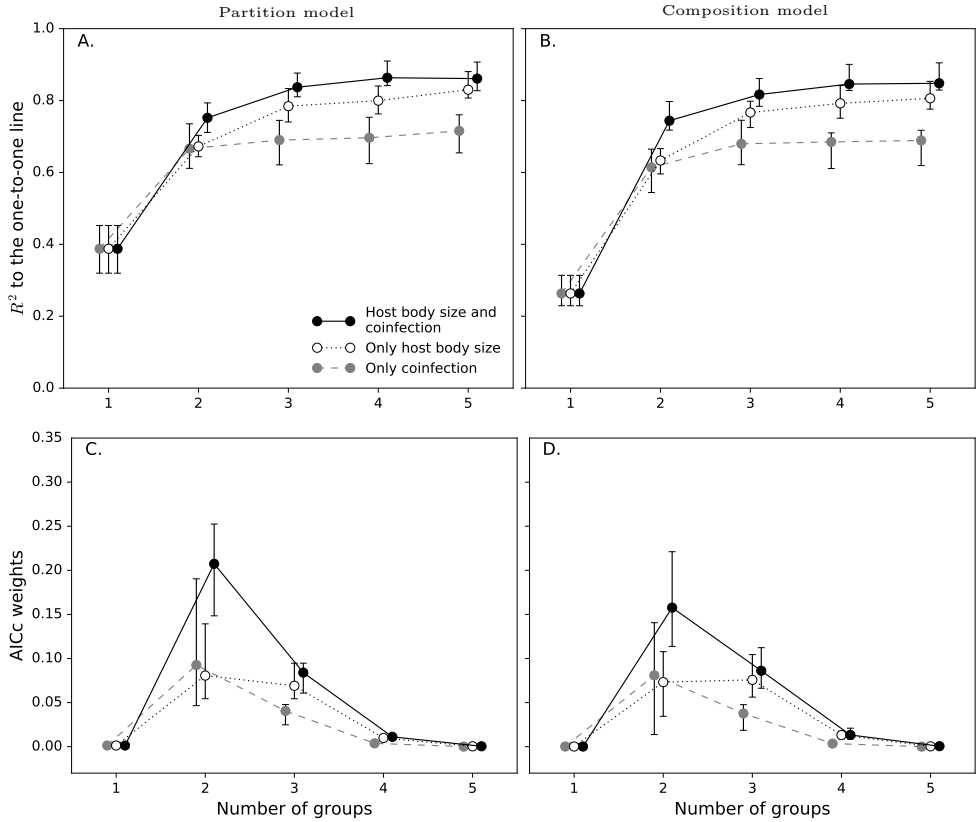


Figure 1.7: Plots show the effects of adding groups of heterogeneity to the constraint-based null models on two metrics: R^2 and AICc weights. Three heterogeneity models with different predictor variables were considered: only host body size, only coinfection with other trematodes, and both body size and coinfection with other trematodes. **A.** and **B.** show how the median R^2 of the 124 overaggregated distributions changes for the partition model (A.) and the composition model (B.) when heterogeneity in body size and and/or coinfection with other trematodes was considered. The points represent the median R^2 and the error bars give the approximate 95% confidence interval around the median. **C.** and **D.** show the median AICc weights for all 124 distributions for a given group size and heterogeneity model. For example, for the partition model every observed host-parasite distribution had 13 candidate models: the no-heterogeneity model (1 group) and three heterogeneity models times four groupings (2 groups, 3 groups, 4 groups, 5 groups). The AICc weights were calculated for these 13 models for a single distribution and then the median AICc weights for a given model was computed for all 124 overaggregated distributions. The error bars give the approximate 95% confidence intervals around these medians. These AICc weights are only comparing within a constraint-based model and are not comparing the partition model (C.) to the composition model (D.).

1.A From feasible sets to the Poisson distribution

The Poisson distribution is the canonical null hypothesis for host-parasite distributions. While there are many ways to obtain a Poisson distribution, the most common derivation in parasite ecology is via a death-immigration process (Anderson and Gordon, 1982). However, one can also obtain a Poisson distribution using the feasible set approach as described below.

As discussed in the main text, the shape of all observed host-parasite distributions are inherently constrained by the total number of parasites in a sample P and the total number of hosts in a sample H. These are hard constraints on the system (Haegeman and Etienne, 2010). Assume now that both hosts and parasites are labeled. For parasites, this means that a given host can have the same parasite intensity (e.g. 1 parasite) in different ways depending on the identity of the individual parasites (e.g. the host is infected with parasite individual A or infected with parasite individual B). For example, given $P = 3$ parasites and $H = 3$ hosts the possible macrostates for the feasible set constrained by $P = 3$ and $H = 3$ are: $\{\{3, 0, 0\}, \{2, 1, 0\}, \{1, 1, 1\}\}$. Given labeled hosts and labeled parasites the probability of seeing the macrostate $\{3, 0, 0\}$ is 3 / 27, the probability of seeing (2, 1, 0) is 18 / 27, and the probability of seeing (1, 1, 1) is 6 / 27. The probabilities of seeing a single host with $x = 0, 1, 2,$ or 3 parasites are $p(0) = 8/27$, $p(1) = 12/27$, $p(2) = 6/27$ and $p(3) = 1/27$.

More generally, given H labeled hosts and P labeled parasites the total number of possible configurations with labeled parasites is $M = H^P$. Given a macrostate m from the feasible set defined by H and P, there are $\mu_m = b_m * \binom{P}{p_1, p_2, \ldots, p_H}$ ways to observed this macrostate with labeled parasites and labeled hosts. b_m is number of ways that macrostate can be realized given unlabeled parasites and labeled hosts (see Methods in

main text), $\binom{P}{p_1,p_2,\ldots,p_H}$ is a multinomial coefficient, and p_i is number of parasites in host i. Taken together, the weight on any particular macrostate m given labeled hosts and labeled parasites is μ_m/M.

Haegeman and Etienne (2010) showed that the probability of observing any particular configuration X of labeled parasites is

$$X \sim \text{Multinomial}(P, p_1 = \frac{1}{H}, p_2 = \frac{1}{H}, \ldots, p_H = \frac{1}{H}) \tag{1.2}$$

From this result, it is easy to see that the probability of a single host having $x = 0, \ldots, P$ parasites is given by a binomial distribution such that

$$x \sim \text{Binomial}(P, \frac{1}{H}) \tag{1.3}$$

Now as H increases (via a larger sample size of hosts) P will also increase (because you are finding more parasites). Therefore, when H is reasonably large and P is reasonably large, the binomial distribution will be well-approximated by a Poisson distribution based on the standard relationship between the Binomial and the Poisson (Zillio and He, 2010).

1.B The central tendency of a feasible set

The central tendency of a feasible set could be found in many different ways, including identifying the mean, median, or mode of the feasible set. While there are advantages and disadvantages to each of these measures of central tendency, we chose to use the vector of marginal medians as our measure for central tendency (Niinimaa and Oja, 2006). Take a feasible set F where F is a matrix with H columns and N rows, where H is the number of hosts and N is the number of macrostates in the feasible set (or

sampled macrostates in the sampled feasible set, Figure 1.S2). The marginal median is computed by taking the median of each of the H columns.

We chose this measure for two reasons. First, because we are dealing with finite samples from the feasible set, we found that the marginal medians were less variable with sample size than the marginal modes (as used by Locey and White, 2013). That being said, using the marginal modes or medians had no effects on our conclusions and tended to yield very similar predictions (e.g. Figure 1.S2). Moreover, the marginal medians were also less sensitive to skew than the mean of the feasible set. Second, we found that the using the marginal median for the composition model yield nearly identical predictions as the predicted rank abundance distribution from the analytical solution for composition model (equation 1 in the main text; Figure 1.S3). This provided further evidence that the marginal medians were a good measure of central tendency for this study.

1.C Goodness-of-fit tests for constraint-based null models

While the R^2 statistic described in the main text provides an easy-to-understand metric of how well a constraint-based model predicts the observed data, it does not account for fact that one of the constraint-based models could have "generated" the data, but the theory and the data might have a low R^2 simply due to sampling error (Xiao et al., 2015*b*). Therefore, we also explored two additional goodness-of-fit tests to determine whether the observed host-parasite distribution differed from a constraint-based null model.

The first test we used was a two-sample Anderson-Darling test which compares the observed RAD and predicted RAD and determines whether these two RADs come from the same distribution (10,000 bootstraps with 'kSamples' R package; Scholz and Zhu, 2015). The null hypothesis of this test is that the two distributions come from the same distribution and is rejected if $p < \alpha$ where α is the Type 1 error rate. Here we set $\alpha = 0.05$.

Using this test, we found that 93% percent of observed host-parasite distributions were not significantly different than the partition model and 90% were not significantly different than the composition model. In contrast, 58% were not significantly different than the Poisson distribution. Fig. 1.S3 shows the proportion of distributions not rejected by the constraint-based models for each host-parasite combination in this study. Across host-parasite combinations, the proportion not rejected by the Anderson-Darling test is largely above 80%. This is a similar pattern to Fig. 1.S2 showing the median R^2 values for each host parasite combination.

While the non-parametric two-sample Anderson-Darling test has more power than other goodness-of-fit tests such as the Kolmogorov-Smirnov test (Engmann and Cousineau, 2011), non-parametric tests inherently lack power compared to parametric alternatives. Therefore, we also implemented a parametric bootstrap test in which we did the following.

1. We computed the log-likelihood of seeing a given observed host-parasite distribution based on the probability mass function $p(x|p, H)$ of seeing a single host with $x = 0, \ldots, P$ parasites for the partition model and the composition model. For the composition model, we used in the analytical formula given in equation 1 in the manuscript. For the partition model, we approximated the probability

mass function by randomly drawing 500 macrostates and computing $p(x|P,H) = \sum_{m \in \hat{F}} p(x|m,P,H)p(m|P,H)$ where m is a macrostate in the sampled feasible set $\hat{F}$.

2. For each observed host-parasite distribution with P and H, we then sampled 500 host-parasite distributions of length H by drawing from $p(x|P,H)$ from the corresponding constraint-based model.

3. For each of these 500 sampled host-parasite distributions, we computed the log-likelihood of seeing the simulated data under the model. This gave us a distribution of 500 log-likelihoods.

4. We then compared the observed log-likelihood to the distribution of sampled log-likelihoods. If the observed log-likelihood did not fall between the the 0.025 and 0.975 quantiles on the sampled log-likelihoods, we rejected that the observed host-parasite distribution was "generated" by the constraint-based model (Rominger and Merow, 2016).

Using this parametric bootstrap approach we found that 99% of the observed host-parasite distributions were not significantly different than the partition model and 96% were not significantly different than the composition model, similar to the values we saw for the Anderson-Darling test. In contrast, only 5% were not significantly different than the Poisson model using this parametric test. This test provides even further support that the observed-host parasite distributions are highly constrained by P and H.

1.D Extending the constraint-based models to include parasite-induced host mortality

Ribeiroia has well-documented negative effects on amphibian survival in the lab and in the field (Johnson, 1999; Johnson et al., 2012). Therefore, we might expect that incorporating the effect of *Ribeiroia*-induced host mortality as an additional constraint on a predicted host-*Ribeiroia* distribution will improve the overall fit of a given constraint-based model to the observed parasite distribution.

To account for *Ribeiroia*-induced mortality, we use the data from the laboratory infection experiments described in Johnson (1999) and Johnson et al. (2012) to estimate an intensity-dependent survival curve for *Pseudacris regilla* infected with *Ribeiroia*. We use a standard logistic survival curve given by

$$\text{logit}(s(x)) = a + bx \tag{1.4}$$

where logit is the logistic function, $s(x)$ is the probability of amphibian survival given a *Ribeiroia* intensity of x, b is the effect of *Ribeiroia* intensity on the log-odds of amphibian survival, and a is the "threshold" at which the host begins to experience parasite-induced mortality. Using a generalized linear model with a binomial response and a logistic link, we estimated the parameters of the *P. regilla*-*Ribeiroia* survival curve to be $a = 1.67$ and $b = -0.05$. See the file `manuscript_analysis_parasite_mortality.py` or Figure 2 in the main text for the data used to fit this GLM.

Using this estimated survival curve, we used a Metropolis-Hastings algorithm to draw a weighted feasible set that accounted for the additional constraint of parasite-induced host mortality. We did this using the following algorithm:

1. Calculate the total number of *Ribeiroia* parasites P and *Pseudacris* hosts H in given empirical host-parasite distribution.

2. Draw an initial candidate macrostate with P and H using the algorithms provided by Locey and McGlinn (2013).

3. For the candidate macrostate, calculate the likelihood of observing this macrostate given the host-survival curve described above (Equation 1.4). To do this, we assumed that each host in a macrostate was independent and calculated the likelihood of observing the macrostate by multiplying the probabilities of observing each host with a given parasite abundance under the estimated survival curve. The assumption of independence is conditional on observing the macrostate, not deriving the macrostate where each host is inherently non-independent given that the total number of parasites in the system is fixed.

4. Propose a new macrostate and calculate its likelihood from equation 1.4. The proposal distribution for drawing a new macrostate is symmetric due to fact that the basic partition model assumes that each macrostate is equally likely (Locey and White, 2013).

5. Take the ratio, r, of the proposed likelihood over the candidate likelihood. If r is greater than 1, accept the proposed macrostate. Otherwise, accept the proposed macrostate with probability r and accept the candidate macrostate with probability $1 - r$.

6. Set the accepted macrostate as your new candidate macrostate and repeat steps 3-5 a large number of times. Discard the first half of the iterations as warm-up/burn-in samples.

The remaining samples give the weighted feasible set with the additional constraint of parasite-induced mortality.

To sample from the composition model with parasite-induced host mortality we used the same procedure described above, but in addition to assigning each proposed macrostate a likelihood based on the survival function, we also assign each proposed macrostate a likelihood based on composition weighting described in the main text. This amounts to multiplying the two likelihoods. As above, the likelihood ratio of the proposed macrostate and the candidate macrostate determines whether to accept or reject the proposed macrostate.

We could also sample from a binomial distribution with a parasite-induced mortality constraint. This is useful as the binomial distribution is the finite equivalent of a Poisson distribution in this case (see Appendix A), which is the standard null model used in parasite ecology (Anderson and Gordon, 1982). We could do this by changing our proposal distribution to a multinomial distribution where the probability of any one of the H hosts encountering a parasite is $1/H$ (Haegeman and Etienne, 2010). The multinomial distribution from which we propose a new configuration X is then given by

$$X \sim \text{Multinomial}(P, p_1 = \frac{1}{H}, p_2 = \frac{1}{H}, \ldots, p_H = \frac{1}{H}) \tag{1.5}$$

We can draw proposal configurations from this multinomial model and, as described above, assign them a likelihood based on both 1) their likelihood given by the estimated survival function and 2) their probability under the multinomial model. Then we accept or reject our proposed configuration based on the ratio of the likelihoods for the proposed and candidate configuration times the probability ratio of the candidate and proposed configurations under the multinomial model. This is the additional weighting on the acceptance ratio imposed by the Metropolis-Hastings algorithm. In summary, this is a

long-winded way of saying that if the survival function likelihood is 1 and there is no effect of parasite mortality, the algorithm will sample from a multinomial distribution whose predicted rank abundance distribution is equal to the predicted RAD from a binomial distribution with P parasites and H hosts. These algorithms are implemented and tested in the accompanying code.

We applied the algorithms and *P. regilla-Ribeiroia* survival curve to all 133 *P. regilla-Ribeiroia* distributions in the dataset. For each constraint-based model with parasite mortality, we ran the Metropolis-Hastings algorithm for 2000 iterations, discarding the first 1000 iterations as warm-up/burn-in samples. We ran this analysis multiple times from different random starting points to ensure the chains were converging to the same stationary distribution. In general, visual inspection of the trace plots of the mortality-constrained feasible set chains showed consistent convergence, good mixing, and generally had acceptance rates above 50%. This high acceptance rate was expected as these chains were designed to have an acceptance rate of 1 (i.e. sampling from the unconstrained distribution) if parasite-induced mortality was not important. The chains of the constrained maximum entropy model had an average acceptance rate of 0.34 and the majority of the chains showed good mixing. However, 9 of the 133 constrained composition model chains had acceptance rates of less than 10% and showed high autocorrelation between samples. Both excluding the distributions resulting from these chains from the analysis and running the chains for longer had no effect on the conclusions we drew about parasite-induced mortality improving the fit of the maximum entropy model. Moreover, these chains were not problematic in the constrained partition model for which we also concluded that parasite-induced host mortality improved the fit of the constraint-based model. Because these under-sampled chains did not affect our inference or conclusions, we included the distributions resulting from these chains in the analysis presented in the

main paper.

1.E Randomization test for heterogeneity

To test whether the host-parasite distribution predicted by the heterogeneity model was better than a host-parasite distribution generated by a randomly grouping hosts, we randomly permuted hosts with their corresponding parasite intensities into groups with the same number of hosts H_j as predicted by the regression tree analysis on a given host-parasite distribution. We then calculated the total number of parasites in each group j and used the procedure described in Figure 3 in the main text to generate the predicted RAD for these randomly permuted mixture distributions. More specifically, this is equivalent to finding the rank abundance distribution for the mixture model $g(x) = \sum_{j=1}^{G} \frac{H_i}{H} p(x|P_{j,rand}, H_j)$ where $P_{j,rand}$ indicates that the total number of parasites in each group j varies with each random permutation. We repeated this randomization 200 times for every empirical host-parasite distribution for both the partition model and composition model. This generated 400 permuted host-parasite distributions (200 using the partition model and 200 using the composition model) for every observed host-parasite distribution. We could then use these permuted samples to determine whether the host groupings produced by the regression tree analysis improved the fit of the constraint-based models to the empirical distributions more than we would expect by randomly permuting hosts into groups. If the mixture model from the regression tree provided a significantly better fit to the host-parasite distribution then randomly grouping hosts, we expected its R^2 value to be significantly higher than the upper bound of the 95% quantile of the R^2 values from the randomly generated groupings.

1.F The levels of aggregation and predictions of constraint-based null models for specific amphibian host-trematode parasite distributions

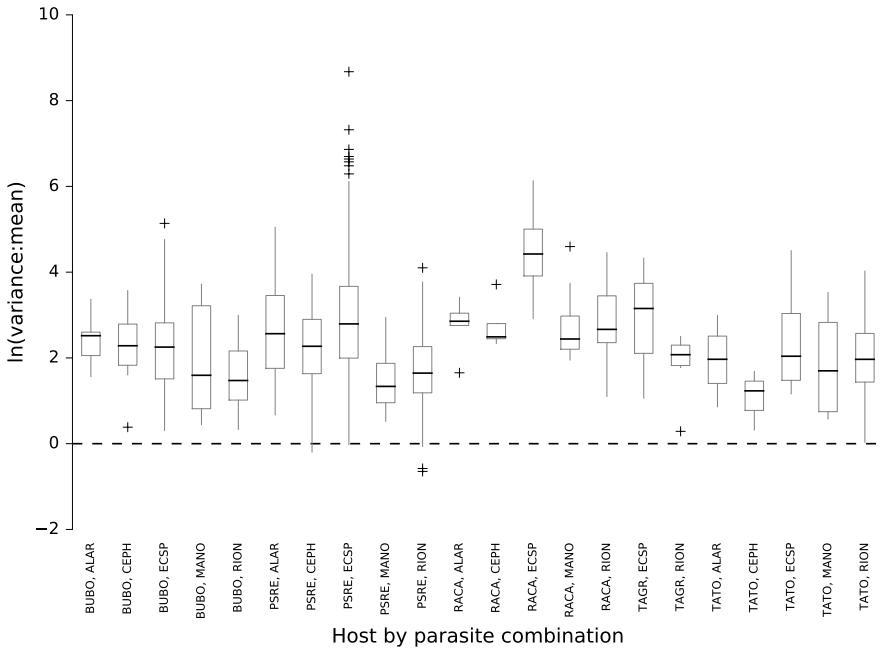


Figure 1.S1: Boxplots of the ln(variance to mean ratio) for each host by parasite combination for all 842 host-parasite distributions used in this analysis. The black dashed line indicates where the ln(variance to mean ratio) is zero, consistent with an unaggregated Poisson distribution. As expected, nearly all of the host-parasite distributions have a ln(variance to mean ratio) greater than 0, indicating an aggregated distribution.

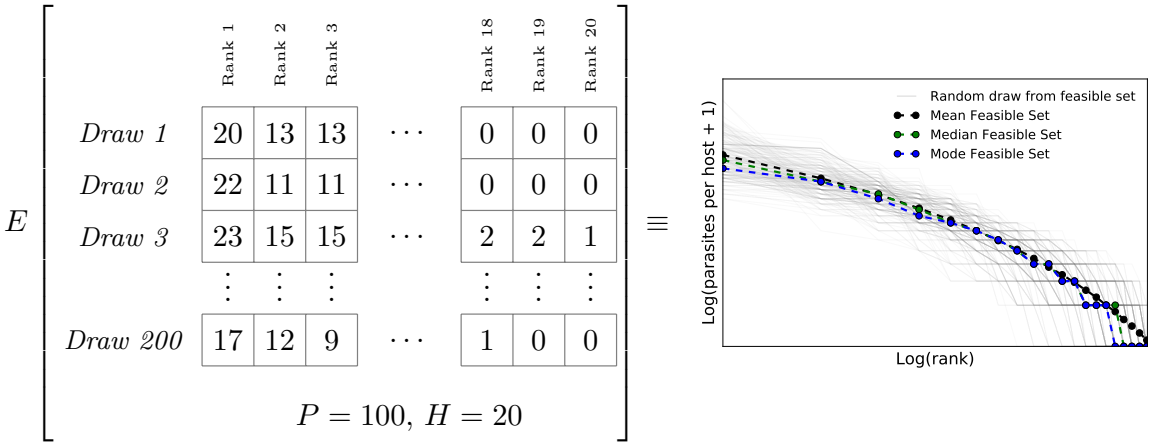


Figure 1.S2: Given a host-parasite system with $P = 100$ parasites and $H = 20$ hosts, the feasible set can be approximated by drawing some number of random macrostates from the full feasible set (200 in this example) and ranking the hosts in each drawn configuration where the host with the most parasites has a rank of 1 and the host with the fewest individuals has a rank of H. The plot shows the graphical representation of this procedure where each gray line is a sampled configuration from the feasible set and the dashed lines are different measures of the center of the sampled feasible set.

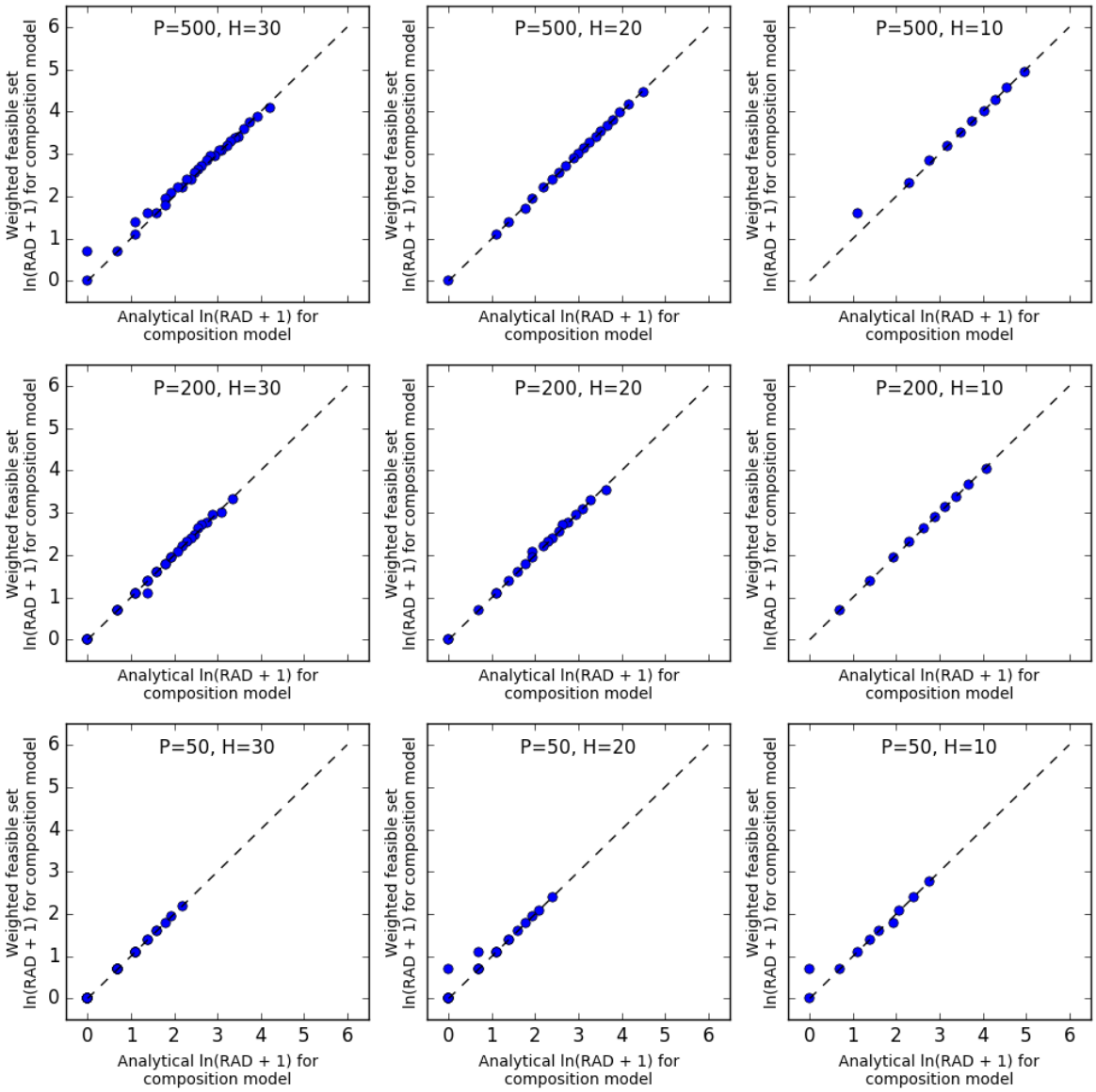


Figure 1.S3: Comparing rank abundance distributions (RAD) from the analytical solution to the composition model (equation 1 in the main text) and the rank abundance distribution (RAD) predicted from the marginal medians of the weighted feasible set of the composition model. Each plot shows a different combination of P and H. Each blue point is a particular host in the RAD with some ln(parasite abundance + 1). The black dashed line is the one to one line. Other than some occasional sampling error, these two methods of predicting the RAD give nearly identical results.

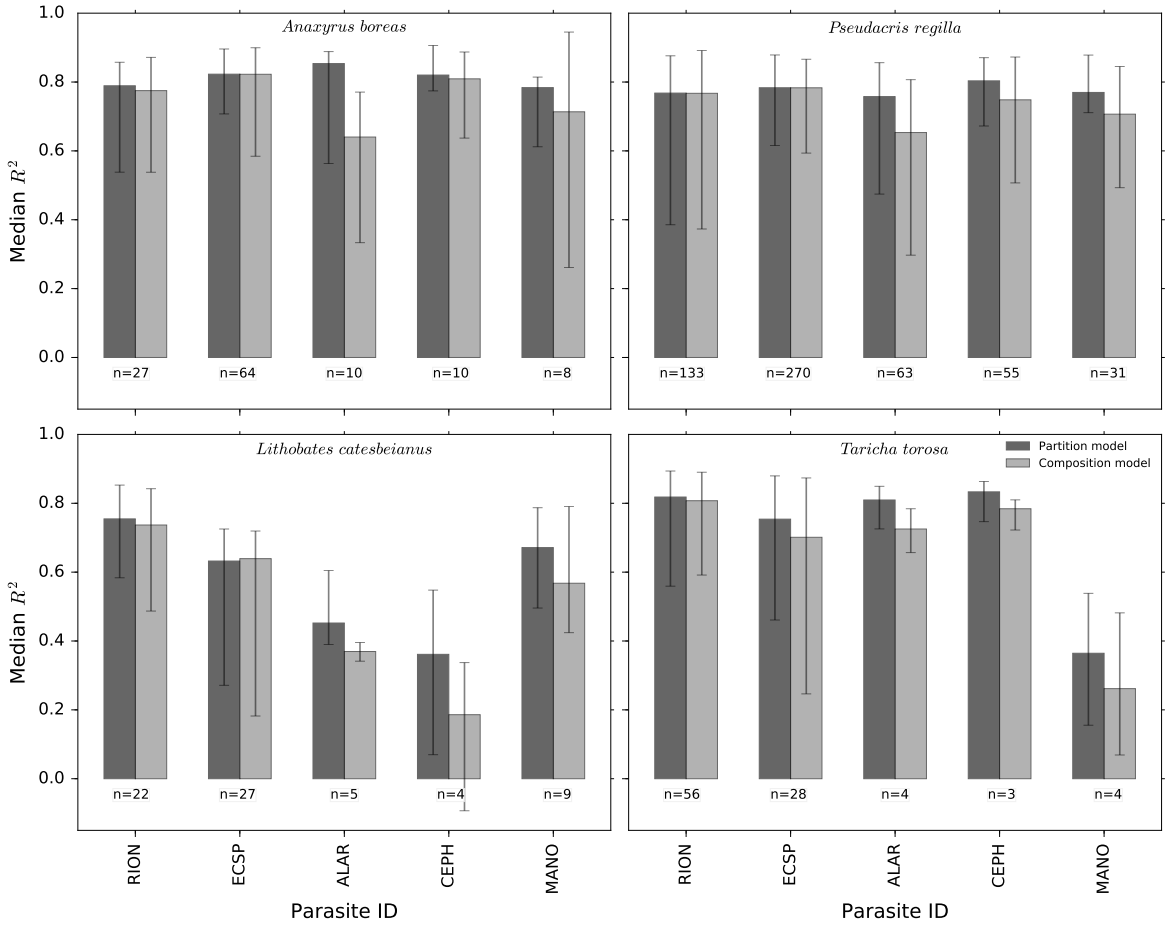


Figure 1.S4: The height of each bar gives the median R^2 about the 1:1 line comparing the observed and predicted rank abundance distributions for various host-parasite combinations. The error bars give the first and third quartiles of the distribution of R^2 values for each distribution of a species host-parasite distribution. The number of distributions for each host-parasite combination that were used to compute this median R^2 are shown in the figure. The x-axis gives the 5 trematode parasites examined in this analysis: *Ribeiroia ondatrae* (RION), *Echinostoma* sp. (ECSP), *Alaria* sp. (ALAR), *Cephalogonimus* sp. (CEPH), and *Manodistomum* sp. (MANO). *Taricha granulosa* is not shown in this plot as it was never infected with ALAR, CEPH or MANO.

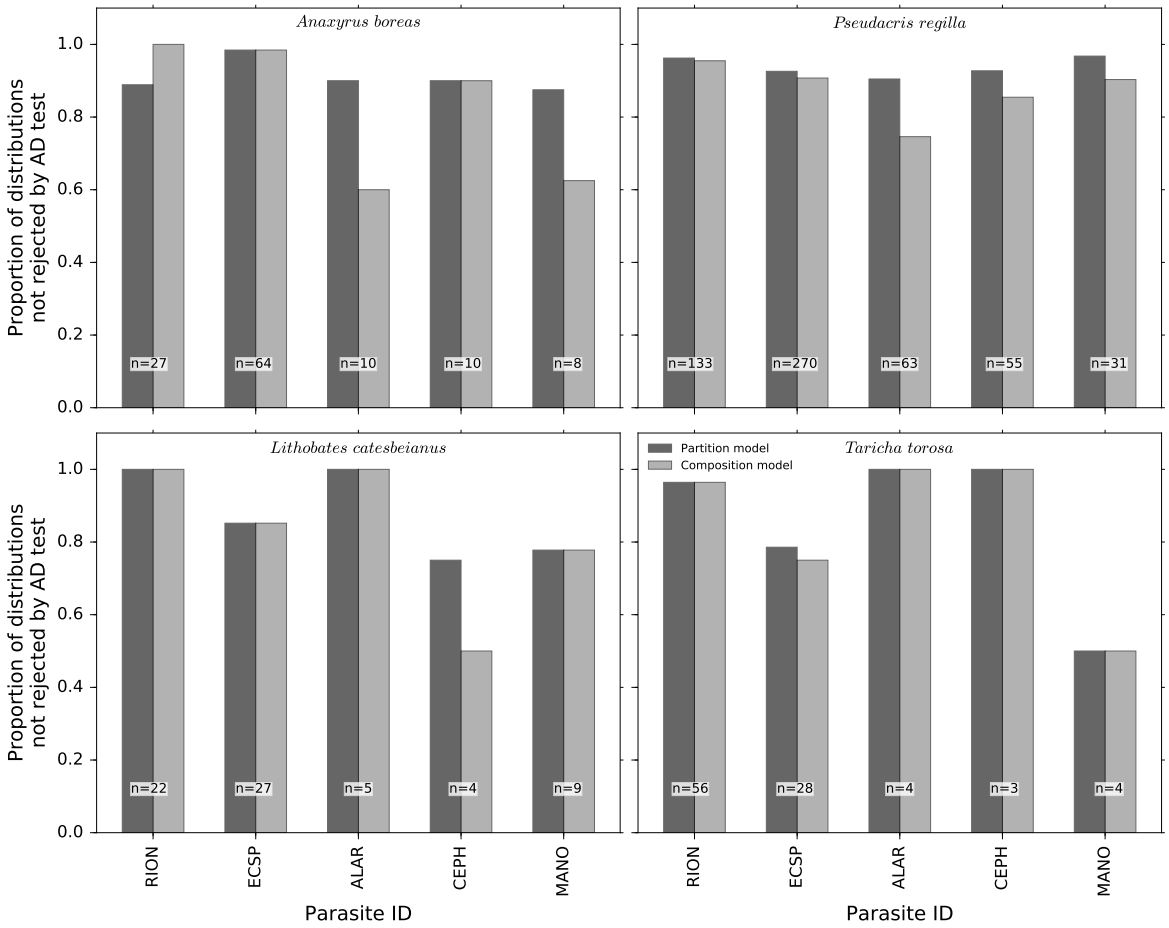


Figure 1.S5: The proportion of predicted host-parasite distributions that were not rejected by the Anderson-Darling test ($\alpha = 0.05$) when compared to the partition model and the composition model. The number of distributions tested for any given host-parasite combination are also displayed on the figure. The x-axis gives the 5 trematode parasites examined in this analysis: *Ribeiroia ondatrae* (RION), *Echinostoma* sp. (ECSP), *Alaria* sp. (ALAR), *Cephalogonimus* sp. (CEPH), and *Manodistomum* sp. (MANO). *Taricha granulosa* is not shown in this plot as it was never infected with ALAR, CEPH or MANO.

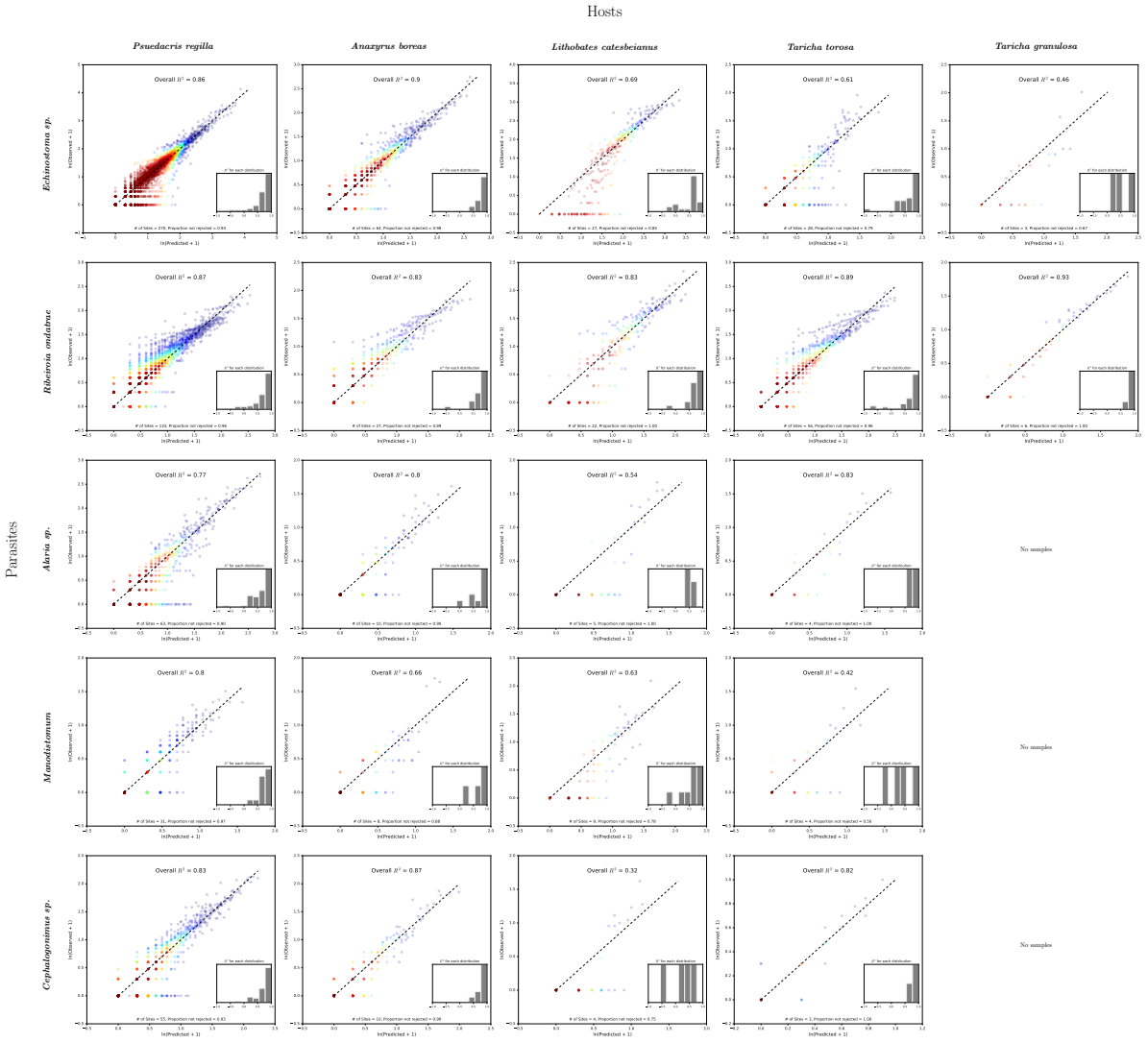


Figure 1.S6: Comparison of observed and predicted host-parasite distributions from the partition model for each host-parasite combination. In each subplot, the x-axis gives the model ln(predicted parasite abundance + 1) and the y-axis gives the ln(observed parasite abundance + 1). Each point gives a particular host's predicted and observed parasite abundance. The color of the points represent the density of points in that region. "Hotter" colors mean there are more points in that region while "cooler" colors mean there are less points in that region. The dashed black line gives the 1:1 line, along which we would expect the points to fall if the model was a perfect fit. The overall R^2 measures the proportion of variance in the data described by the model. The text at the bottom of the plot gives the total number of distributions that were tested and the proportion of them that were not rejected by an Anderson-Darling test at $\alpha = 0.05$. Finally, the histogram in the lower right hand corner of each plot gives the histogram of the R^2 values for each of the host-parasite distributions for a given host-parasite combination.

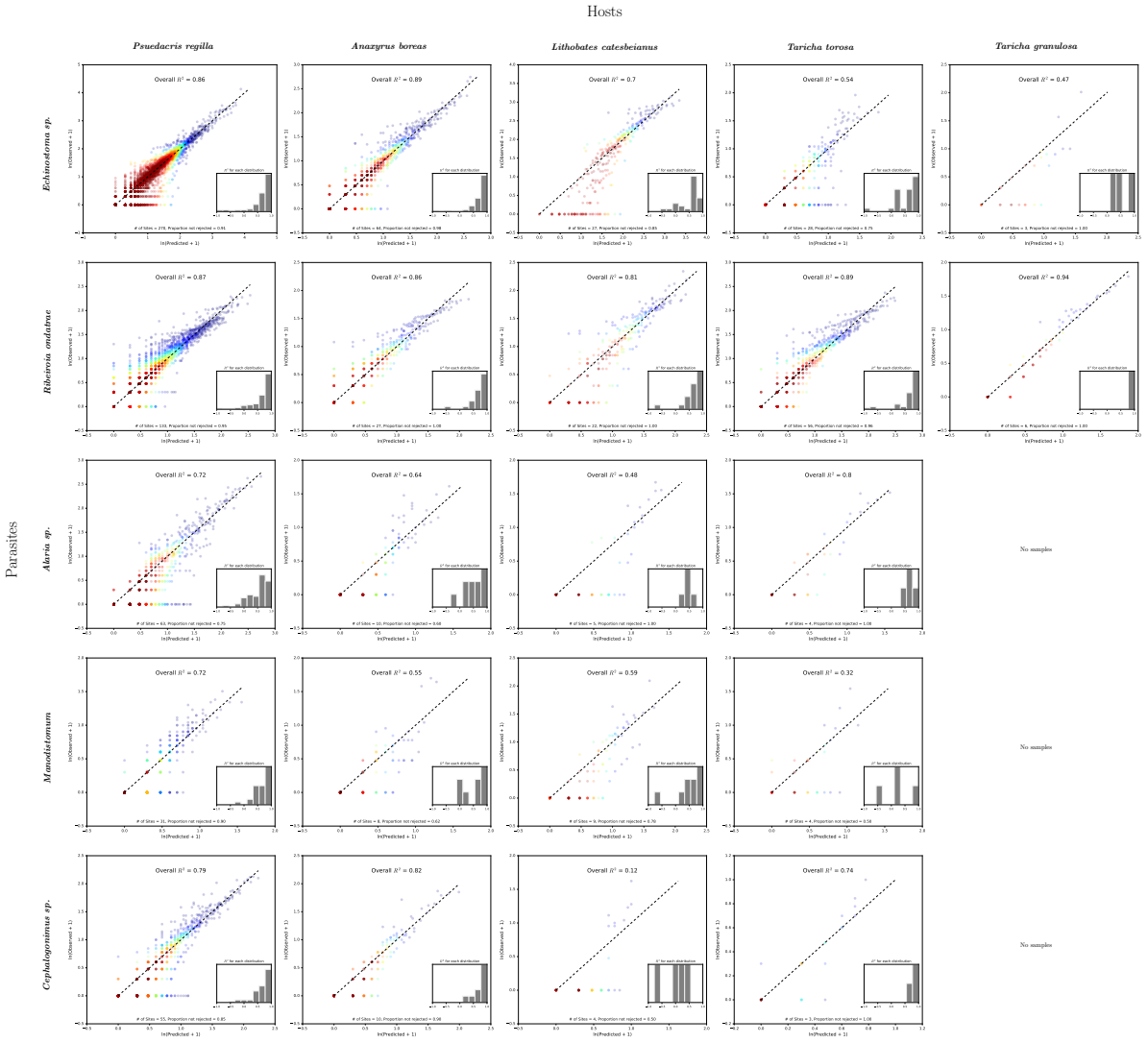


Figure 1.S7: Comparison of observed and predicted host-parasite distributions from the composition model for each host-parasite combination. In each subplot, the x-axis gives the model ln(predicted parasite abundance + 1) and the y-axis gives the ln(observed parasite abundance + 1). Each point gives a particular host's predicted and observed parasite abundance. The color of the points represent the density of points in that region. "Hotter" colors mean there are more points in that region while "cooler" colors mean there are less points in that region. The dashed black line gives the 1:1 line, along which we would expect the points to fall if the model was a perfect fit. The overall R^2 measures the proportion of variance in the data described by the model. The text at the bottom of the plot gives the total number of distributions that were tested and the proportion of them that were not rejected by an Anderson-Darling test at $\alpha = 0.05$. Finally, the histogram in the lower right hand corner of each plot gives the histogram of the R^2 values for each of the host-parasite distributions for a given host-parasite combination.

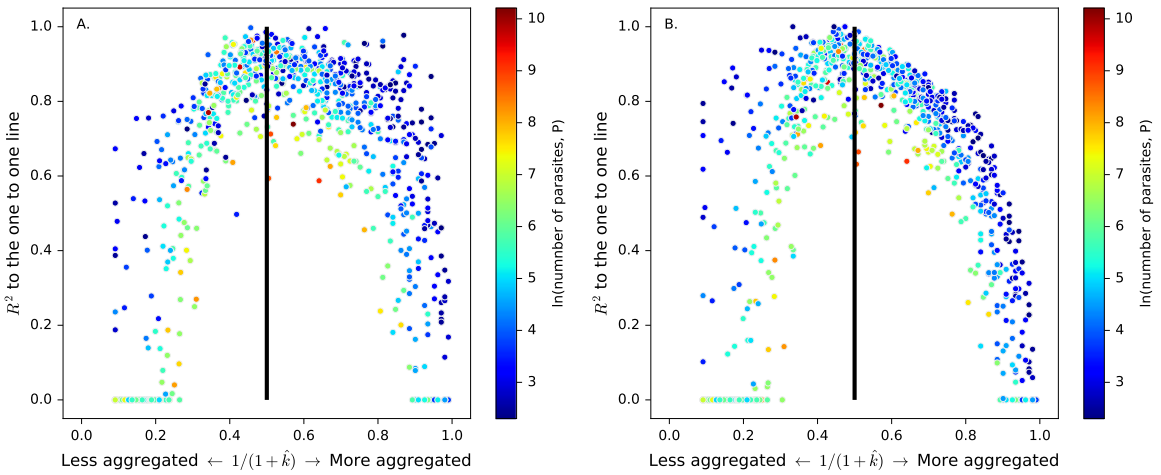


Figure 1.S8: **A.** gives the partition model R^2 values plotted against the maximum likelihood estimates of the k parameter from the negative binomial distribution ($\hat{k}$) for all 842 host-parasite distributions used in this study. Each point is a host-parasite distribution and the color of each point indicates how many parasites (P) were in that distribution. The bold vertical line indicates where $\hat{k} = 1$ or $1/(1+\hat{k}) = 0.5$, which corresponds to the composition model. $\hat{k}$ was transformed for visual clarity. Also note that if R^2 to the one to one line was less than 0, which is valid based on equation 1 in the main text, we set it equal to 0 for visual clarity. **B.** provides the same information, except that the R^2 values are from the composition model. Note that because of error in the estimate of $\hat{k}$ due to finite sampling, $\hat{k}$ may not equal one and R^2 can still be well above 0.5. Also note that the partition model, when compared to the composition model, tends to yield in higher R^2 values for distributions that are more aggregated than $\hat{k} = 1$ or $1/(1+\hat{k}) = 0.5$.

Chapter 2

Detecting and quantifying parasite-induced host mortality from intensity data: method comparisons and limitations

2.1 Abstract

Parasites can significantly impact animal populations by changing host behavior, reproduction and survival. Detecting and quantifying these impacts is critical for understanding disease dynamics and managing wild animal populations. However, for wild hosts infected with macroparasites, it is notoriously difficult to quantify the fatal parasite load and number of animals that have died due to disease. When ethical or logistical constraints prohibit experimental determination of these values, examination of parasite intensity and distribution data may offer an alternative solution. In this study we introduce a novel method for using intensity data to detect and quantify parasite-induced mortality in wildlife populations. We use simulations to show that this method is more reliable than previously proposed methods while providing quantitative estimates of parasite-induced mortality from empirical data that are consistent with previously published qualitative estimates. However, this method, and all techniques that estimate parasite-induced mortality from intensity data alone, have several important assumptions that must be scrutinized before applying them to real-world data. Given that these assumptions are met, our method is a new exploratory tool that can help inform more rigorous studies of parasite-induced host mortality.

2.2 Introduction

Infectious agents can impact animal populations by changing population dynamics and stability (Dobson and Hudson, 1992; Tompkins et al., 2002), altering predator-prey interactions (Joly and Messier, 2004), and even causing species' decline and extinction (De Castro and Bolker, 2005; McCallum, 2012). Accurately estimating the impact of

these infectious agents in wildlife is critical to understanding what regulates host and parasite populations, making predictions about disease transmission, and managing disease outbreaks (Langwig et al., 2015). The impact of pathogens, such as rabies (Coyne et al., 1989), bovine tuberculosis (Cox et al., 2005), and rinderpest (Tillé et al., 1991), are typically modeled based on the presence or absence of disease, such that host survival is not generally considered to be a function of the number of infectious agents present within the host. In contrast, models of macroparasites generally assume that pathology increases with parasite burden and host survival probability must be treated as a function of infection intensity (Anderson and May, 1978). Helminths exhibiting this intensity-dependent pathology have significant impacts on human health (Brooker et al., 2004), domestic livestock economics (Roeber et al., 2013), and wildlife survival (Kirk, 2003; Logiudice, 2003). While it is generally assumed that some fraction of wild host populations succumb to parasitic infection, it is notoriously difficult to actually quantify parasite-induced host mortality (PIHM) in wild animal populations because it is difficult to observe the dead or dying hosts most impacted by parasitism (McCallum, 2000).

Ideally, parasite-induced host mortality is quantified by experimentally infecting and tracking individual hosts in the wild population; however, for logistical and ethical reasons this method is rarely feasible (McCallum, 2000). Snapshot data of parasite intensities across multiple hosts is much easier to collect and has often been used to identify the presence of PIHM (Crofton, 1971; Lester, 1977, 1984; Lanciani and Boyett, 1989; Royce and Rossignol, 1990; Ferguson et al., 2011) and to quantify the relationship between infection intensity and host mortality (Adjei et al., 1986).

Crofton (1971) first proposed that PIHM could be identified from parasite intensity data by comparing the observed parasite distribution in sampled hosts to the distribution

predicted in the absence of parasite-induced mortality. This method assumes that, prior to host mortality, infection intensity in the host population follows a negative binomial distribution and the tail of the distribution is truncated as intensity dependent pathology removes the most heavily infected hosts. Assuming mortality occurs only in heavily infected hosts, evidence of this parasite-induced mortality should then be detectable by iteratively fitting a negative binomial distribution to hosts with lower and lower parasite intensities, and comparing these truncated predicted distributions to the corresponding truncated observed parasite data (Fig. 2.1, see Section 2.A for additional detail).

While the Crofton Method detects the presence of PIHM, it makes no attempt to quantify the relationship between infection intensity and host survival probability; information that is necessary for estimating parasite impacts on host populations (Anderson and May, 1978; Tompkins et al., 2002). Adjei et al. (1986) suggested that this relationship could be calculated by first using the Crofton Method to estimate the pre-mortality parasite distribution and then using this distribution to calculate the probability of host survival with increasing parasite intensity. To do this, Adjei et al. (1986) modeled host survival as a logistic function and then used a generalized linear model (GLM) to estimate the parameters of the host survival function (see Section 2.B for a technical description of the Adjei Method). Although this method can predict the host survival function, it has several technical drawbacks. When mean infection intensity is high or sample sizes are small the observed intensity data must be subjectively binned into intensity ranges in order to fit the GLM framework. Furthermore, for the Adjei Method to work, any observed intensity values greater than predicted values must be modified and set equal to the predict values (see Section 2.B for details); a questionable act of data manipulation. These manipulations may introduce bias, reduce the precision and limit the power of this method to detect and quantify parasite-induced host mortality.

After 30 years, and despite clear limitations (McCallum, 2000), these methods (particularly the Crofton Method) are still discussed among parasitologists and are the primary techniques for examining population-level impacts of parasitism using parasite intensity data. In these methods, PIHM can only be identified by visually examining plots of the pre-mortality parameters predicted by the Crofton Method and determining whether they show a "kink" over a range of truncation values (Fig. 2.1B; Lester, 1984; Ferguson et al., 2011). These qualitative criteria makes it difficult to compare PIHM between studies and a more rigorous and quantitative method is needed to both detect and quantify host mortality. The survival function given by the Adjei Method may be used to do this; however, it requires manipulating the original data and its accuracy remains untested.

In this study, we propose a novel method for detecting and quantifying PIHM that ameliorates many of the aforementioned deficiencies of the previous methods. Our method does not require data alteration, is highly generalizable, and uses standard statistical techniques to quantitatively determine whether PIHM is occurring in a system. We use simulations to compare our method with the Adjei Method to test the ability of both to (1) detect occurrence of PIHM and (2) estimate the host survival function. We then apply both methods to real datasets previously used in PIHM analyses and compare the results. Finally, we discuss the limitations of inferring PIHM from intensity data and how these methods fit in modern quantitative parasitology.

2.3 Materials and methods

2.3.1 A novel, likelihood-based method for estimating PIHM

Our method (henceforth the Likelihood Method) begins with the same assumptions as the Adjei Method: namely that infection has occurred and hosts with fatal parasite loads have died prior to the population sampling. As discussed by Adjei et al., this is not necessarily unrealistic as some parasite infections occur primarily in younger hosts with parasite-induced mortality occurring soon after infection (e.g. Schotthoefer et al., 2003; Johnson and McKenzie, 2008).

The Likelihood Method then assumes that prior to mortality the parasite distribution can be described by the distribution $g(x;\phi)$, which specifies the probability of a host having x parasites before mortality occurs. ϕ is a vector of parameters that describes the shape of this distribution. The probability of a host surviving with x parasites from infection until sampling is given by the host survival function $h(\text{survival};x,\theta)$ where θ specifies any additional parameters needed to define the host survival function.

With these two assumptions, we can define a distribution that gives the probability of having a parasite load of x parasites conditional on host survival, $P(x|\text{survival})$. Using standard rules of conditional probability this distribution can be written as

$$P(x|\text{survival}) = \frac{P(\text{survival}|x) * P(x)}{P(\text{survival})} \tag{2.1}$$

$P(\text{survival}|x)$ is the survival function $h(\text{survival};x,\theta)$, $P(x)$ is the pre-mortality parasite distribution $g(x;\phi)$ and
$P(\text{survival}) = \sum_{x=0}^{\infty} P(\text{survival}|x) * P(x) = \sum_{x=0}^{\infty} h(\text{survival};x,\theta) * g(x;\phi)$. Therefore,

equation 1 can be written as

$$P(x|\text{survival}) = \frac{h(\text{survival}; x, \theta) * g(x; \phi)}{\sum_{x=0}^{\infty} h(\text{survival}; x, \theta) * g(x; \phi)} \tag{2.2}$$

Using this probability distribution, one can then find the parameters θ and ϕ that maximize the likelihood of an observed host-parasite dataset. To estimate the significance of PIHM in a host-parasite system, a likelihood ratio test can be used in which the full model is given by equation 2 and the reduced model is given by the pre-mortality distribution $g(x; \phi)$. If PIHM is not significant in the system, the resulting likelihood ratio statistic should approximately follow a χ^2 distribution with degrees of freedom equal to the number of parameters in the full model with parasite-induced mortality minus the number of parameters in the reduced model without parasite-induced mortality (Bolker, 2008).

The parameterization of equation 2.2 depends on the parasite system of interest. Here, we assume that the pre-mortality parasite distribution $g(x; \phi)$ follows a negative binomial distribution with two parameters mean parasite intensity (μ_p) and aggregation (k_p, where smaller k_p indicates a more aggregated parasite population) before mortality (Crofton, 1971; Anderson and May, 1978; Adjei et al., 1986). A variety of different biological and statistical assumptions can result in an equilibrium parasite distribution that follows a negative binomial distribution (Kendall, 1948; Boswell and Patil, 1970; Calabrese et al., 2011). Furthermore, the negative binomial distribution is an incredibly flexible distribution that fits many host-parasite systems even when the underlying mechanisms determining the empirical distribution are unknown (Shaw et al., 1998).

The function for $h(\text{survival}; x, \theta)$ is also system specific. Many theoretical models of parasite-induced host mortality assume that the parasite-induced death rate of hosts is a linear function of parasite intensity (Anderson and May, 1978; Dobson and Hudson,

1992; Barbour and Pugliese, 2000). In systems where there is truly a linear relationship between infection intensity and survival probability it will be nearly impossible to use intensity data to detect parasite-induced host mortality (Lanciani and Boyett, 1989). However, some systems do exhibit non-linear host survival functions (Benesh, 2011), in which case these methods would be applicable.

To compare the Likelihood Method and the previously proposed Adjei Method, we adopt the non-linear, logistic host-survival function used in the earlier study given by

$$h(\text{survival}; x, a, b) = \frac{\exp\left(a - b\log(x)\right)}{1 + \exp\left(a - b\log(x)\right)} \tag{2.3}$$

Generally, a larger b leads to a more rapid decline in the probability of host survival as parasite intensity increases, with the maximum rate of decline having a value of $b/4$ (Section 2.B). b is in many ways analogous to the pathogenicity parameter (α) in classic macroparasite models that gives the parasite intensity dependent host death rate (Anderson and May, 1978; Isham, 1995). When b is held constant, a larger a allows for hosts to tolerate larger parasite intensities before experiencing parasite-induced mortality. More specifically, for every one unit increase in a the log parasite intensity at which any percent of hosts survive (e.g. 99% of hosts survive) increases by $1/b$ (Section 2.B).

The equation $\exp(a/b)$ can also be used to calculate the parasite LD_{50}, here defined as the infection intensity above which a host has greater than 50% probability of dying. Equation 3 is commonly used in toxicology and has the useful properties of being bounded between 0 and 1 and being differentiable for all x (Collet, 2002). That being said, it is phenomenological and is used simply because it tends to fit survival data. However, given that a goal of these analyses is to compare the Likelihood Method's results to the Adjei Method, it is natural to adopt the same host-survival function to

facilitate comparison. When applying the Likelihood Method to other systems more mechanistic host-survival functions can be used in place of equation 3.

2.3.2 Evaluating the Adjei and Likelihood Methods

Question 1: Can we detect PIHM?

We used statistical power and Type I error to test the ability of the Adjei Method and the Likelihood Method to correctly identify the presence of PIHM on simulated data with known pre-mortality parameters. The power of a method is the probability of correctly detecting PIHM given that it is occurring and the Type I error is the probability of incorrectly identifying PIHM given that it is not occurring. If a method has low Type I error we can be confident that when we detect PIHM it is actually occurring. If one method has higher power for detecting PIHM than another, we will need to sample fewer hosts to detect PIHM.

Consistent with the model assumption that parasite infection, host mortality, and population sampling are temporally separate events, we first created a pre-mortality host population by drawing N_p randomly infected hosts from a negative binomial distribution with parameters μ_p and k_p. This represents a host population that has become infected but not yet experienced parasite-induced mortality (Adjei et al., 1986). In the Adjei Method and Crofton Method, N_p is a necessary parameter defined as the number of hosts in the population before parasite-induced mortality. More accurately, N_p is the number of hosts that would have been sampled had parasite-induced host mortality not occurred. This parameter is not necessary when using the Likelihood Method because, unlike the Adjei Method and Crofton Method which estimate parasite-induced mortality using absolute numbers of hosts, the Likelihood Method estimates parasite-induced mortality

using probabilities. However, to compare the results of the Likelihood Method with the Adjei Method, we specified a value for N_p for all simulations.

We next chose values of a and b for the host survival function and calculated the probability of survival for all N_p hosts using equation 3. Then, to simulate the period in which hosts died due to infection, for each host we drew a random number from a uniform distribution between 0 and 1 and if the calculated host survival probability was less than this random number, the host experienced parasite-induced mortality. The surviving individuals represent the post-mortality hosts that would be sampled in the field.

We then used these simulated pre-mortality and post-mortality datasets to test the ability of both methods to correctly determine whether or not PIHM was occurring when the parameters N_p, μ_p and k_p were known. Although the parameters N_p, μ_p, and k_p are always unknown in real systems, a method that fails under these ideal simulation conditions with known parameters will certainly also fail when these values must be estimated from empirical data. In practice, for the Adjei Method, N_p, μ_p, and k_p are estimated using the Crofton Method (Adjei et al., 1986), while μ_p and k_p in the Likelihood Method can be estimated jointly with a and b or via the Crofton Method.

We compared the two methods using three different mean parasite intensity values (μ_p = 10, 50, 100) and three different host survival functions (gradual, moderate, and steep decreases in the host survival with increasing parasite intensity, Fig. 2.2A). For a given μ_p, each survival function had the same LD_{50} ([$\mu_p = 10$, $LD_{50} = 7.39$], [$\mu_p = 50$, $LD_{50} = 35.57$], [$\mu_p = 100$, $LD_{50} = 77.3$]), but different values of a and b. We examined each μ_p-survival function pair at three levels of parasite aggregation, $k_p = 0.1$, 0.5, and 1 — realistic values of parasite aggregation in natural populations (Shaw et al., 1998). For each of these 27 parameter combinations we simulated 150 datasets and tested the

probability of each method correctly identifying PIHM in the post-mortality dataset (power) and incorrectly identifying PIHM in the pre-mortality dataset (Type I error). For each method, we used a likelihood ratio test to determine whether the full model with PIHM provided a significantly better fit than the reduced model without PIHM at significance level of 0.05. We also examined the impact of sample size by simulating each parameter for pre-mortality sample sizes of $N_p = [50, 100, 200, 300, 400, 500]$. Wild host populations were assumed to be sampled after PIHM has occurred, thus we calculated the sample size in the power simulations as the average number of surviving hosts over all 150 simulations for each parameter combination. The distribution of surviving hosts over the 150 simulations was generally symmetrical and the standard deviation was small compared to the mean (maximum coefficient of variation was approximately 0.06 across all parameter combinations), suggesting that the mean number of surviving hosts was an adequate summary statistic of the number of hosts sampled post-mortality.

We then tested the ability of the Likelihood Method to correctly identify PIHM under the more realistic condition of unknown pre-mortality parameters. Based on the first set of simulations, we excluded the Adjei Method and only examined the power of the Likelihood Method under "best-case" scenario parameter values, setting $\mu_p = 10$ and $k = 1$ because PIHM is most detectable when parasites are less clumped and mean intensity is low. We examined the impact of survival function shape and sample size on the Likelihood Method's ability to identify PIHM when the pre-mortality parameters μ_p and k_p and the survival function parameters a and b needed to be estimated. We performed 500 simulations over a range of different samples sizes for gradual, moderate, and steep survival functions, following the simulation procedure described above.

Question 2: Can we estimate properties of the host survival function?

In the previous section we compared the ability of the Adjei Method and the Likelihood Method to provide a "yes" or "no" answer for whether or not PIHM was occurring in a system. In this section we compared the ability of the Adjei Method and the Likelihood Method to estimate properties of the survival function such as the parameters a, b and LD_{50}. Using the same simulation procedure and parameter combinations described above, we simulated 150 datasets, estimated a, b, and LD_{50} and calculated the standardized bias and precision for these estimates (Walther and Moore, 2005). Because estimating properties of the host survival function requires more information than simply detecting PIHM, we used larger values of N_p for this simulation (N_p = [300, 500, 1000, 2000, 5000, 7500, 10000]). We used the average number of surviving hosts for each set of 150 simulated datasets as our measure of sample size. Although both a and b are necessary to estimate LD_{50}, the two parameters showed similar patterns of bias and precision so we only show the results for a.

2.3.3 Application to real data

We tested the ability of the Adjei Method and the Likelihood Method to identify PIHM in six host-parasite datasets given in Crofton (1971) and four datasets given in Adjei et al. (1986) (Table 1). Crofton analyzed infection patterns in the snail *Gammarus pulex* infected with the acanthocephalan *Polmorphus minutus*. Adjei et al. analyzed males and females of two species of lizard fish *Saurida tumbil* and *S. undosquamis* that were infected by the cestode *Callitetrarhynchus gracilis*.

In both earlier studies, the authors reported PIHM in some of the datasets and we tested whether the Adjei Method and/or the Likelihood Method also predicted PIHM. For the six datasets from Crofton (1971), we used the general conclusions of the author

and truncated the data at four parasites, applied the Crofton Method to estimate the pre-mortality distribution, and then ran the Likelihood Method and Adjei Method using these pre-mortality parameters. For the Adjei et al. (1986) datasets, we followed the same procedure as the authors and first truncated the data at two parasites and then fit the Crofton Method for the female fish of both species. Then, following the Adjei et al.'s methods, we parameterized the male pre-mortality distributions for each species with the results from the females. Finally, we applied the Adjei Method and the Likelihood Method to determine whether or not PIHM was significant for these species and compared our results to those given by the authors. All code for the analyses is provided in Section 2.C.

2.4 Results

2.4.1 Question 1: Detecting presence of PIHM

The power of the Adjei Method to detect PIHM in a system was close to unity for larger sample sizes and tended to decrease as sample size decreased for all survival functions (Fig. 2.2C; Fig. 2.S1-2.S3). The Likelihood Method had a power close to unity for all parameter combinations and sample sizes considered. With gradual survival functions, the power of the Likelihood Method decreased slightly for small samples sizes (Fig. 2.2C, Fig. 2.S1-2.S3).

The Adjei Method had highly inflated Type I error rates (i.e. falsely detected PIHM) for all parameter combinations that we considered (Fig. 2.2B; Fig. 2.S1-2.S3). This method also showed the unintuitive pattern of decreasing Type I error rate with decreasing sample size. This occurred because, at small samples sizes, intensity data must

be binned before the Adjei Method can be used (Section 2.B). In contrast, the Likelihood Method showed a Type I error rate at or near the pre-set level of 0.05 for all parameter combinations and sample sizes considered (Fig. 2.2B; Fig. 2.S1-2.S3).

When all parameters were jointly estimated, the Likelihood Method showed highly context-dependent results even when detecting PIHM under the best-case scenario of $\mu_p = 10$ and $k_p = 1$. For steep survival curves, PIHM could be detected with a power of greater than 0.8 from a sample of less than 100 hosts (Fig. 2.3). However, for moderate survival functions over 400 hosts had to be sampled to achieve the same power and for gradual survival functions, no tested sample size ever achieved a power greater than 0.8 (Fig. 2.3).

2.4.2 Question 2: Estimating the LD_{50} and survival function

The Likelihood Method gave asymptotically unbiased estimates of the LD_{50} for all combinations of parameters examined in this study (Fig. 2.4, Fig. 2.S4-2.S6). Even for the smallest sample sizes we considered, the Likelihood Method's estimate of LD_{50} was largely unbiased, with small biases occurring for gradual host survival functions. The precision of the Likelihood Method's LD_{50} estimates decreased (increasing coefficient of variation) as sample size decreased for all parameter combinations we examined (Fig. 2.4, Fig. 2.S4-2.S6).

The Adjei Method produced biased estimates of the LD_{50} across nearly all parameter combinations, tending to underestimate the true value of the parameter (Fig. 2.4, Fig. 2.S4-2.S6). For $\mu_p = 10$, the LD_{50} estimates from the Adjei Method were largely unbiased for large samples sizes, but as μ_p increased, the Adjei Method produced biased estimates of LD_{50} across all sample sizes, with bias increasing as sample size decreased

(Fig. 2.4, Fig. 2.S4-2.S6). The LD_{50} estimates from the Adjei Method also showed large decreases in precision with the steepest survival function across all values of μ_p (Fig. 2.4, Fig. 2.S4-2.S6).

In terms of the host survival function, the Likelihood Method gave unbiased estimates of survival function parameter a when sample sizes were large, however as sample size decreased these estimates became severely biased (Fig. 2.4, Fig. 2.S7-2.S9) The Adjei Method produced biased estimates of the host survival function across all sample sizes, with consistently greater bias for steeper survival functions and higher mean parasite loads. (Fig. 2.4, Fig. 2.S7-2.S9).

2.4.3 Application to real data

The previous authors qualitatively detected PIHM in 7 of the 10 datasets considered (Table 1). The Likelihood Method parameterized from the pre-mortality parameters of the Crofton Method detected significant PIHM in 6 of these 7 datasets at a significance level of 0.05. The only dataset in which the Likelihood Method did not detect a significant effect of PIHM was the Adjei dataset for female *S. tumbil*. For this dataset there was a marginally significant effect of PIHM ($\chi^2_{df=2} = 5.34; p = 0.069$). The Adjei Method detected PIHM in 9 of the 10 datasets (Table 1), consistent with our simulation results showing that the Adjei Method has a high Type I error rate. Moreover, the Adjei Method estimates of the LD_{50} were quite variable for the Crofton data, consistent with our simulation results that the Adjei Method LD_{50} estimates could be imprecise for sample sizes of less than 1000 hosts (Fig. 2.S4-2.S6).

2.5 Discussion

Our likelihood-based method to estimate parasite-induced host mortality from observed parasite intensity data is a significant improvement over the previous methods. In simulations, it had greater power for detecting PIHM over a wider range of parameter values and also exhibited fewer false detection events (Type I errors) in both simulations and when applied to published datasets previously used in PIHM analyses. The Likelihood Method was also generally less biased and more precise when quantifying parasite-induced mortality via the host survival function for the parameters we considered. The superior performance of the Likelihood Method over the Adjei Method can be attributed to its fewer parameters, its lack of unnecessary data alteration, and its applicability across a variety of different parameter combinations. In short, the Likelihood Method is a better method for detecting and quantifying PIHM than the previously proposed Adjei Method.

Although superior to the Adjei Method, the Likelihood Method still cannot be applied to all real datasets. For host-parasite systems where host mortality occurs as a steep, non-linear function of parasite intensity only 75 hosts must be sampled to have an 80% power in detecting PIHM. However, as the maximum slope of the survival function decreases and the function becomes somewhat linear, hundreds, or possibly thousands of hosts would have to be sampled to achieve the same result. This is consistent with previous studies which illustrate the difficulty of detecting PIHM from linear host survival functions (Lanciani and Boyett, 1989). While it may be feasible to sample several hundred invertebrates or small fish, even the smallest sample sizes are completely unfeasible for many vertebrates, particularly the species of conservation concern where addressing the impact of parasitism would be most important. An even larger sample

size would be required to identify PIHM when parasites are highly aggregated, mean infection intensity is high, or parasite prevalence is low, all of which are common in many parasitic helminths. Moreover, while linear functions make PIHM undetectable, at the other extreme, steep, non-linear survival curves produce severely biased estimates of the survival function. Given the interaction between all of these different factors, the Likelihood Method is probably limited to detecting PIHM in systems where greater than 100 hosts can be collected, parasites are common and only moderately aggregated, and substantial host mortality occurs at relatively low parasite intensity.

While we have improved on the existing methods for quantifying PIHM from parasite intensity data, all such methods require several fundamental assumptions. Nearly all current methods derive from Crofton (1971) (but see Ferguson et al., 2011) and assume that, prior to any PIHM, parasites are distributed in the host population following a negative binomial distribution. But, it is fundamentally impossible to know what the pre-mortality parasite distribution was in a wild host population and it is widely recognized that different processes can lead to a variety of parasite distributions in hosts (Anderson and Gordon, 1982; Duerr et al., 2003). However, the negative binomial is extremely flexible and there is substantial empirical and theoretical evidence to support the assumption that, prior to any PIHM, parasite distributions can be fit by a negative binomial distribution (Shaw and Dobson, 1995; Shaw et al., 1998; Wilson et al., 2002).

It is important to note that the flexibility of the negative binomial distribution may also reduce our ability to detect PIHM. If a negative binomial can be fit to the observed post-mortality parasite distribution then, regardless of how lethal the parasite was, it will be impossible to detect PIHM because there is no need for a more complex model. Many observed parasite distributions are well-fit by the negative binomial distribution (Shaw et al., 1998), suggesting that systems where these methods are applicable without

any *a priori* knowledge may be uncommon. However, if one has *a priori* knowledge about some aspect of the pre-mortality distribution (e.g. assumes/knows the value of k_p, Ferguson et al., 2011), then the Likelihood Method could be applicable even if the the post-mortality distribution was well-fit by a negative binomial.

If one has evidence that the pre-mortality is not negative binomial, the generality of our method easily allows another distribution to be specified for $g(x, \phi)$. For example, one could use the resulting stationary host-parasite distribution from a stochastic host-parasite model without parasite-induced host mortality (Anderson and Gordon, 1982) to specify the form of $g(x, \phi)$ and then apply the techniques discussed in this paper to detect PIHM. The general requirement for the Likelihood Method to detect PIHM in a stochastic host-parasite process is that the stationary distribution of the process with mortality is significantly different than the stationary distribution without mortality. It is widely recognized that parasite-induced host mortality decreases the aggregation of host-parasite distributions relative to those without mortality (Barbour and Pugliese, 2000), suggesting that the Likelihood Method could be generally applicable to host-parasite systems that follow the assumptions of many stochastic host-parasite models. This is an intriguing area for further research.

If the Likelihood Method is applicable and the truncation of the negative binomial distribution is detected, one must be aware that the truncation pattern may be caused by other processes such as within host density dependence, age dependent variation in host resistance and/or heterogeneous infection rates (Anderson and Gordon, 1982; Rousset et al., 1996; McCallum, 2000). This means that in the event that PIHM is detected, it may actually not be the result of PIHM. Moreover, if host mortality depends on parasite intensity and additional variables (e.g. host sex, host size), failure to identify these important confounding variables could significantly affect the ability of these methods

to correctly identify PIHM. However, both of these issues – inferring process from pattern and confounding variables – are well-recognized limitations of most statistical inference and are addressed via judicious model specification and selection (Seber and Lee, 2003).

As suggested by Lester (1984) these methods for estimating PIHM can provide preliminary insight into whether or not PIHM is worth further exploration. However, we stress that these methods are an exploratory tool for assessing the role of PIHM in a system, and potential users should critically evaluate whether they think they have a large enough sample size and an appropriate host survival function/post-mortality distribution for the methods developed in this paper to be applicable. Even if they are applicable, inferring PIHM from distributional data is no substitute for field experiments and an in depth understanding of the natural history of the host-parasite system under consideration.

2.6 Acknowledgments

We thank the Briggs Lab group, Kuris/Lafferty Lab group, Theoretical Ecology group at University of California, Santa Barbara (United States), and two anonymous reviewers for helpful feedback. MW was supported by National Science Foundation Graduate Research Fellowship (Grant No. DGE 1144085) and the University of California Regents (United States). CB was supported by National Institute of Health (United States) grant 1R01GM109499 from the Ecology of Infectious Disease program.

References

Adjei, E. L., A. Barnes, and R. J. G. Lester. 1986. A method for estimating possible parasite-related host mortality, illustrated using data from *Callitetrarhynchus gracilis* (Cestoda: Trypanorhyncha) in lizardfish (*Saurida* spp.). Parasitology. **92**:227–243.

Anderson, R. M., and D. M. Gordon. 1982. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. Parasitology. **85**:373–398.

Anderson, R. M., and R. M. May. 1978. Regulation and stability of host-parasite interactions: I. Regulatory processes. J. Anim. Ecol. **47**:219–247.

Barbour, A. D., and A. Pugliese. 2000. On the variance-to-mean ratio in models of parasite distributions. Adv. Appl. Probab. **32**:701–719.

Benesh, D. P. 2011. Intensity-dependent host mortality: what can it tell us about larval growth strategies in complex life cycle helminths? Parasitology. **138**:913–25.

Bolker, B. M. 2008. Ecological Models and Data in R. Princeton University Press, Princeton, New Jersey.

Boswell, M. T., and G. P. Patil, 1970. Chance mechanisms generating the negative binomial distributions. *in* G. P. Patil, editor. Random Counts in Scientific Work Vol. 1. Pennsylvania State University Press.

Brooker, S., J. Bethony, and P. J. Hotez. 2004. Human hookworm infection in the 21st century. Adv. Parasit. **58**:197–288.

Calabrese, J. M., J. L. Brunner, and R. S. Ostfeld. 2011. Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model. PlOS One **6**:e29215.

Collet, D., 2002. Bioassay and some other applications. Chapter 4, pages 103–118 *in* Modelling Binary Data. Chapman & Hall, London.

Cox, D. R., C. A. Donnelly, F. J. Bourne, G. Gettinby, J. P. McInerney, W. I. Morrison, and R. Woodroffe. 2005. Simple model for tuberculosis in cattle and badgers. P. Natl. Acad. Sci.-Biol. **102**:17588–17593.

Coyne, M. J., G. Smith, and E. McAllister, Fiona. 1989. Mathematic model for the population biology of rabies in raccoons in the mid-Atlantic states. Am. J. Vet. Res. **50**:2148–2154.

Crofton, H. D. 1971. A quantitative approach to parasitism. Parasitology. **62**:179–193.

De Castro, F., and B. Bolker. 2005. Mechanisms of disease-induced extinction. Ecol. Lett. **8**:117–126.

Dobson, A. P., and P. J. Hudson. 1992. Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenuis* in red grouse. II. Population models. J. Anim. Ecol. **61**:487–498.

Duerr, H. P., K. Dietz, and M. Eichner. 2003. On the interpretation of ageintensity profiles and dispersion patterns in parasitological surveys. Parasitology. **126**:87–101.

Ferguson, J. A., W. Koketsu, I. Ninomiya, P. A. Rossignol, K. C. Jacobson, and M. L. Kent. 2011. Mortality of coho salmon (*Oncorhynchus kisutch*) associated with burdens of multiple parasite species. Int. J. Parasitol. **41**:1197–205.

Isham, V. 1995. Stochastic models of host-macroparasite interaction. Ann. Appl. Probab. **5**:720–740.

Johnson, P. T. J., and V. J. McKenzie, 2008. Effects of Environmental Change on Helminth Infections in Amphibians: Exploring the Emergence of *Ribeiroia* and *Echinostoma* Infections in North America. Chapter 11, pages 249–280 *in* B. Fried and R. Toledo, editors. The Biology of Echinostomes: From the Molecule to the Community. Springer-Verlag New York.

Joly, D. O., and F. Messier. 2004. The distribution of *Echinococcus granulosus* in moose: Evidence for parasite-induced vulnerability to predation by wolves? Oecologia **140**:586–590.

Kendall, D. G. 1948. On the generalized "birth-and-death" processes. Ann. Math. Stat. **19**:1–15.

Kirk, R. S. 2003. The impact of *Anguillicola crassus* on European eels. Fisheries Manag. Ecol. **10**:385–394.

Lanciani, C. A., and J. M. Boyett. 1989. Demonstrating parasitic water mite-induced mortality in natural host populations. Parasitology. **81**:465–475.

Langwig, K. E., J. Voyles, M. Q. Wilber, W. F. Frick, K. A. Murray, B. M. Bolker, J. P. Collins, T. L. Cheng, M. C. Fisher, J. R. Hoyt, D. L. Lindner, H. I. McCallum, R. Puschendorf, E. B. Rosenblum, M. Toothman, C. K. Willis, C. J. Briggs, and A. M. Kilpatrick. 2015. Context-dependent conservation responses to emerging wildlife diseases. Front. Ecol. Environ. **13**:195–202.

Lester, R. J. G. 1977. An estimate of mortality in a population of *Perca flavescens* owing to the trematode *Diplostomum adamsi*. Can. J. Zoolog. **55**:288–292.

Lester, R. J. G. 1984. A review of methods for estimating mortality due to parasites in wild fish populations. Helgolander Meeresun. **37**:53–64.

Logiudice, K. 2003. Trophically transmitted parasites and the conservation of small populations: raccoon roundworm and the imperiled allegheny woodrat. Conserv. Biol. **17**:258–266.

McCallum, H. 2012. Disease and the dynamics of extinction. Philos. T. R. Soc. B. **367**:2828–39.

McCallum, H. I., 2000. Host-pathogen and host-parasite models. Chapter 10, pages 284–312 *in* J. H. Lawton and G. E. Likens, editors. Population Parameters: Estimation for Ecological Models. Blackwell Science Ltd.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models. Second edition. Chapman & Hall, New York.

Roeber, F., A. R. Jex, and R. B. Gasser. 2013. Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for exploring epidemiology and drug resistance - an Australian perspective. Parasites and Vectors **6**:153.

Rousset, F., F. Thomas, T. D. Meeûs, and F. Renaud. 1996. Inference of parasite-induced host mortality from distributions of parasite loads. Ecology **77**:2203–2211.

Royce, L. A., and P. Rossignol. 1990. Epidemiology of honey bee parasites. Parasitol. Today. **6**:348–353.

Schotthoefer, A. M., R. A. Cole, and V. R. Beasley. 2003. Relationship of tadpole stage to location of echinostome cercariae encystment and the consequences for tadpole survival. J. Parasitol. **89**:475–482.

Seber, G. A. F., and A. J. Lee. 2003. Linear Regression Analysis. John Wiley and Sons, Inc, Hoboken, New Jersey.

Shaw, D. J., and A. P. Dobson. 1995. Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. Parasitology. **111**:111–133.

Shaw, D. J., B. T. Grenfell, and A. P. Dobson. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology. **117**:597–610.

Tillé, A., C. Lefèvre, P. P. Pastoret, and E. Thiry. 1991. A mathematical model of rinderpest infection in cattle populations. Epidemiol. Infect. **107**:441–452.

Tompkins, D. M., A. P. Dobson, P. Arneberg, M. Begon, I. M. Cattadori, J. V. Greenman, J. A. P. Heesterbeek, P. J. Hudson, D. Newborn, A. Pugliese, A. P. Rizzoli, R. Rosa, F. Rosso, and K. Wilson, 2002. Parasites and host population dynamics. Chapter 3, pages 45–62 *in* P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heessterbeck, and A. P. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Walther, B. A., and J. L. Moore. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. Ecography. **28**:815–829.

Wilson, K., O. N. Bjoernstad, A. P. Dobson, S. Merler, G. Poglayen, A. F. Read, and A. Skorping, 2002. Heterogeneities in macroparasite infections: patterns and pro-

cesses. Chapter 2, pages 6–44 *in* P. J. Hudson, A. Rizzoli, B. Grenfell, H. Heesterbeek, and A. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

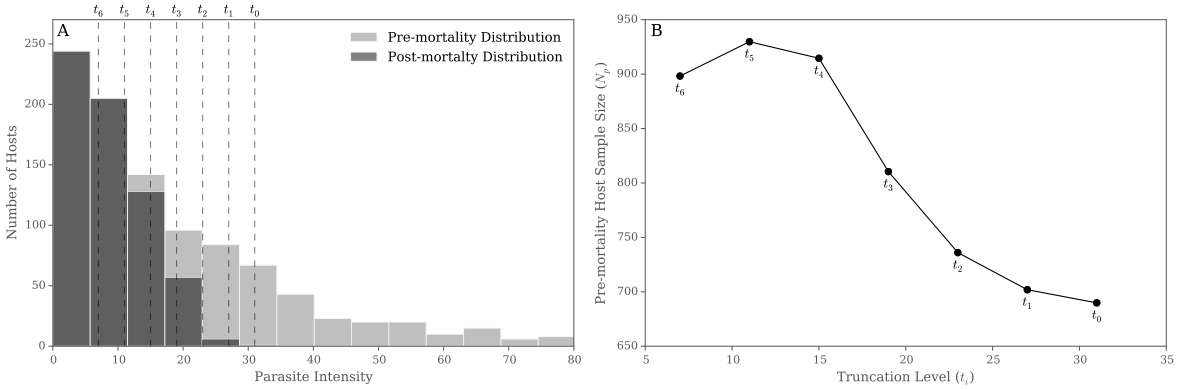


Figure 2.1: A schematic representation of the iterative approach of the Crofton Method. (A) The light gray shows the pre-mortality distribution that the Crofton Method is trying to estimate from the dark grey post-mortality distribution. The Crofton Method proceeds by truncating the post-mortality data at different levels (t_i, e.g. $i = 0, \ldots, 6$) and finding the pre-mortality host population size (N_p), pre–mortality mean parasite intensity (μ_p), and pre-mortality parasite aggregation (k_p) that best fit the truncated data. (B) The parameter N_p is then plotted against the truncation level t_i to determine if a "kink" occurs in the parameter values (Lester, 1984). This "kink" indicates that PIHM is occurring in the system. In the above example, PIHM is occurring in the system as visualized by the distinct "kink" at t_4.

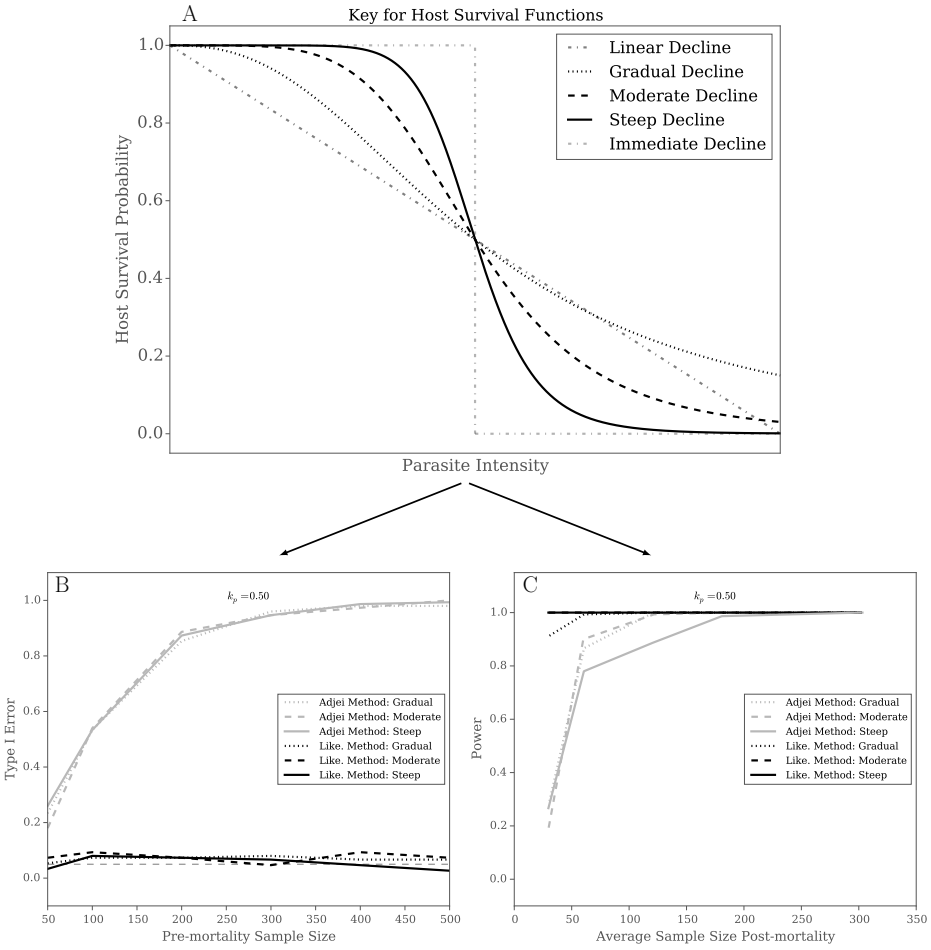


Figure 2.2: The simulation results comparing the power and the Type I error of the Adjei Method and the Likelihood Method across a range of different sample sizes. (A) Five potential shapes for a host-survival functions. In the simulations we used a gradual survival function (dotted line), and moderate survival function (dashed line), and a steep survival function (solid line). The linear and immediate survival functions represent two potential extremes that we do not include in the simulations. For each of these survival functions and the parameter combinations described in the main text, we tested the Type I error and power of the Likelihood (Like.) Method and Adjei Method. (B) Gives the Type I error of each method over a range of pre-mortality sample sizes with a pre-mortality mean parasite intensity (μ_p) of 50 and pre-mortality parasite aggregation (k_p) at 0.5. The red line shows the pre-set significance level of 0.05. (C) Gives the power of each method for detecting PIHM over a range of post-mortality sample sizes for $\mu_p = 50$ and $k_p = 0.5$. In general, the Likelihood Method has higher power and lower Type I error than the Adjei Method. See the Fig. 2.S1-2.S3 for Type I error and power results for all parameter combinations.

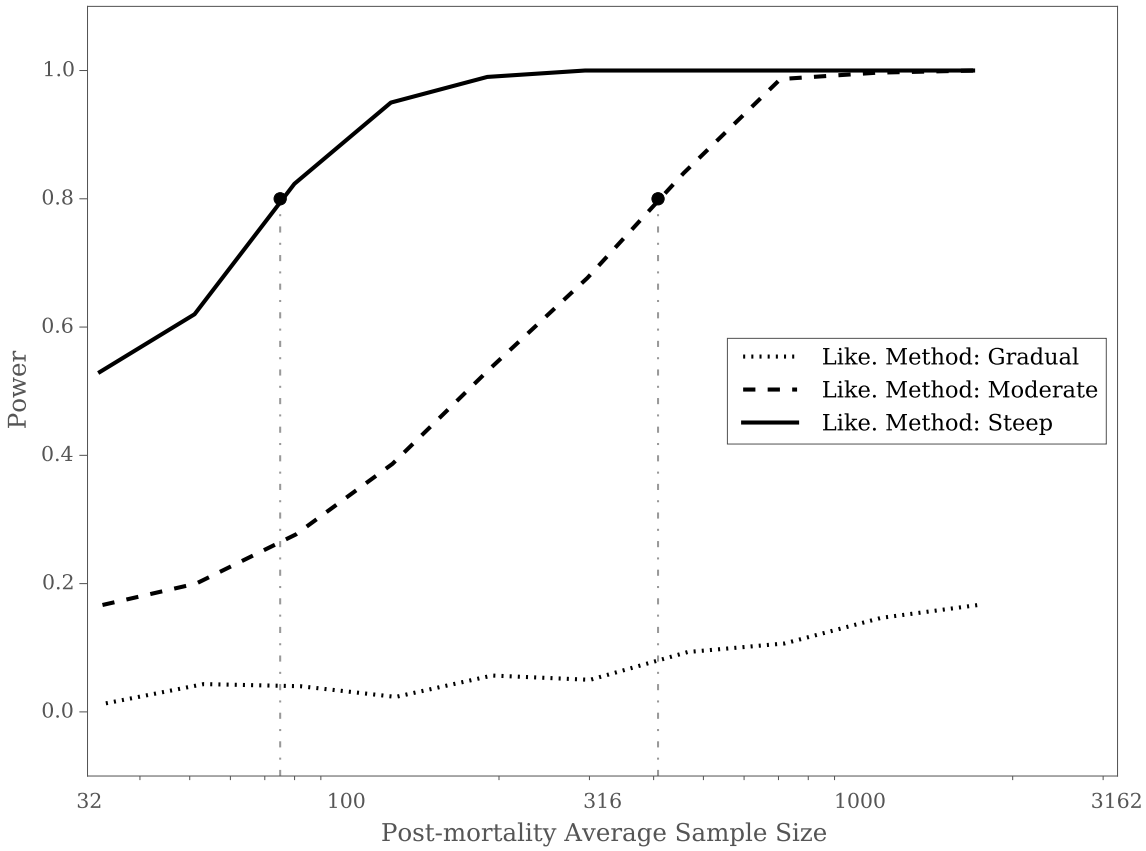


Figure 2.3: The power of the Likelihood Method (Like.) to detect PIHM for gradual, moderate, and steep survival functions when all four parameters μ_p, k_p, a, and b were jointly estimated. The curves were generated from 500 simulations for 10 pre-mortality sample sizes, N_p. The vertical, dotted-dashed lines indicate the sample size at which the power for the Likelihood Method with steep and moderate survival functions is 0.8 (75 hosts for steep functions and 408 for moderate functions). The Likelihood Method with a gradual survival function never has a power above 0.8.

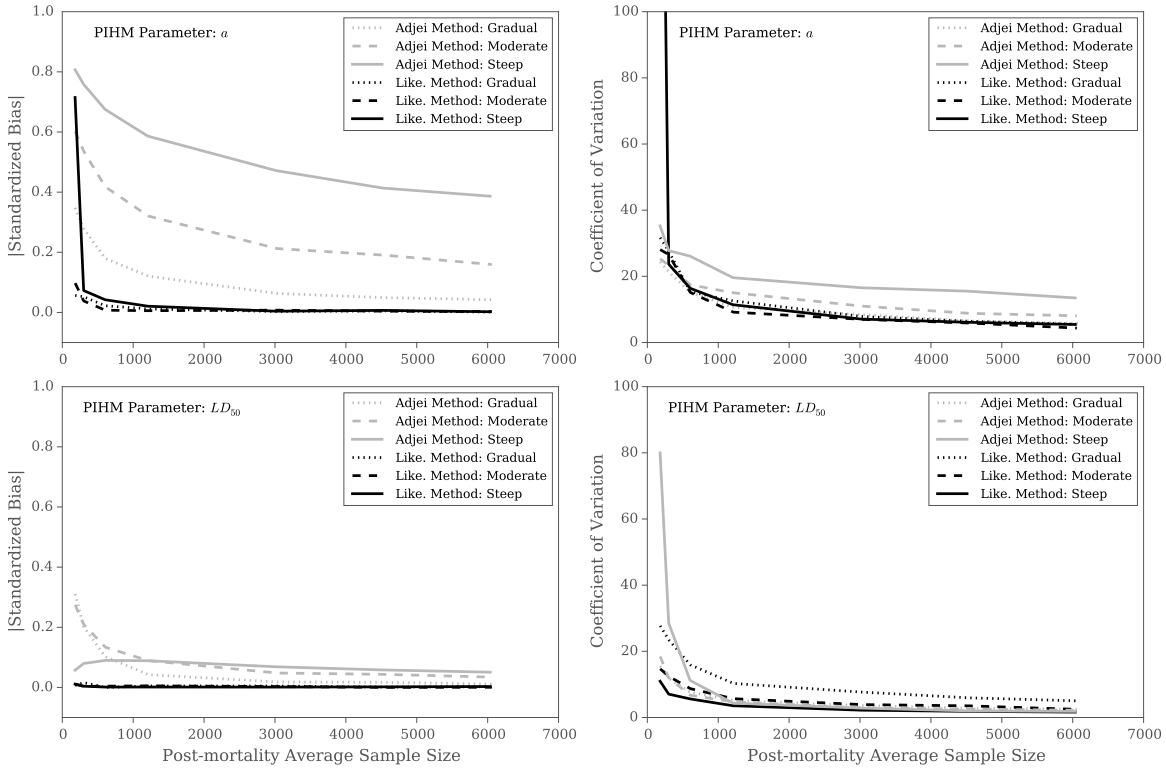


Figure 2.4: Bias and precision (coefficient of variation) for the Likelihood Method (Like.) and Adjei Method estimates of the a parameter and the LD_{50} of the host survival function based on simulated PIHM data over a range of post-mortality sample sizes. As the coefficient of variation increases, precision decreases. The pre-mortality parameters for this simulation were $\mu_p = 50$ and $k_p = 0.5$. The figure shows the simulations for three different host survival functions (gradual, moderate, and steep), each with the same LD_{50}. Bias and precision results of LD_{50} and a for all other parameter combinations can be found in Fig. 2.S4-2.S9.

2.A Implementation of the Crofton Method

The algorithm for fitting the Crofton Method (Crofton, 1971) proceeds as follows. First, obtain a dataset with n hosts where each host has some parasite intensity 0 to p_{max}. Starting with the full dataset, guess a vector of pre-mortality parameters $(N_{p1}, \mu_{p1}, k_{p1})$ where N_{p1} is the total number of hosts before mortality, μ_{p1} is the parasite intensity before mortality, and k_{p1} is the parasite aggregation before mortality. Given these parameters, use a negative binomial distribution to calculate the predicted number of hosts with $0, 1, 2, \ldots p_{max}$ parasites. Compare the expected number of hosts with $0, 1, 2, \ldots p_{max}$ parasites to the observed number hosts with $0, 1, 2, \ldots p_{max}$ parasites and calculate the χ^2-squared statistic associated with the observed and predicted vectors. In reality, one often has to bin the parasite intensity data because all parasite intensities are not represented in the dataset. Continue to guess $(N_{p1}, \mu_{p1}, k_{p1})$ vectors until a set of parameters is found that minimizes the χ^2- squared statistic.

Second, choose a truncation value (t_2) such that $t_2 < p_{max}$. Truncate the data such that $\text{data}_{\text{truncated}} \leq t_2$ and repeat the above iterative procedure to calculate another set of parameters $(N_{p2}, \mu_{p2}, k_{p2})$ that minimizes the χ^2-squared statistic on the truncated data. Choose a new truncated value $t_3 < t_2$ and repeat the first two steps. Continue to truncate the dataset until it only contains hosts with 0, 1, and 2 parasites (or 3 bins). Because the method attempts to estimate three parameters, at least 3 classes are needed for all 3 parameters to be identifiable (Royce and Rossignol, 1990).

Once the iterative procedure has been completed, parasite-induced host mortality is traditionally identified by plotting the different truncation values t_i against the different values of N_{pi} and looking for a distinct "kink" in the resulting plot. Once the "kink" as occurred, the values of N_{pi} will typically remain close to constant as t_i is decreased

further. The "true" pre-mortality parameters N_{pt}, μ_{pt}, and k_{pt} are taken to be at the point where the "kink" occurs.

We provide an implementation and unit tests of the Crofton Method in Section 2.C. Figure 2.S10 visually shows that our implementation of the Crofton Method agrees with results previously published by Crofton (1971).

2.B Implementation of the Adjei Method

The Adjei Method for estimating PIHM has two steps (Adjei et al., 1986). The first step is to estimate the parameters of the pre-mortality host-parasite distribution using the Crofton Method (see Section 2.A). The three parameters estimated are the total number of hosts before mortality N_p, the mean number of parasites per host before mortality μ_p, and the aggregation of parasites before mortality given by the parameter k_p from a negative binomial distribution. When k_p is small, parasites are highly aggregated among hosts and when k_p is large parasites are more evenly distributed across hosts (Wilson et al., 2002). The implementation of the Crofton Method has been discussed at length elsewhere (e.g. Royce and Rossignol, 1990; Lester, 1984, and in Section 2.A) and we provide a tested implementation of the method in Section 2.C.

The second step of the Adjei Method is to make the assumption that infection, host mortality, and sampling occur in that order and are temporally separate (Adjei et al., 1986). Next, Adjei et al. assume that the host survival function follows the logistic form

$$h(\text{survival}; x, a, b) = \frac{\exp\left(a - b\log(x)\right)}{1 + \exp\left(a - b\log(x)\right)} \tag{2.4}$$

where x is the parasite intensity in a given host and a and b are the two parameters of the logistic function. Generally, a larger a allows for hosts to tolerate larger parasite

intensities before experiencing parasite-induced mortality and a larger b leads to a more rapid decline in the probability of host survival as parasite intensity increases. The value $\exp(a/b)$ is referred to as the LD_{50}. Individuals with loads higher than this will have a greater than 50% chance of death.

By taking the first and second derivatives of equation 1, one can easily find that the maximum rate of decline in host survival probability with increasing parasite intensity occurs at the LD_{50} and has a value of $b/4$. This is in many ways analogous to the parasite pathogenicity parameter α given in classic macroparasite models, which specifies the slope of the linear relationship between between host death rate and parasite intensity (Anderson and May, 1978; Isham, 1995). The parameter a is easily interpreted by holding b constant and looking at how a one unit change in a affects the log parasite intensity at which some percentage p of hosts experience mortality. Letting a_1 and a_2 be two different values of a and x_1 and x_2 be two different parasite intensities, a bit of rearranging of equation 1 gives

$$\log \frac{p}{1-p} = a_1 - b \log x_1$$

$$-\log \frac{p}{1-p} = a_2 - b \log x_2$$

$$0 = a_1 - a_2 - b \log x_1 + b \log x_2$$

$$a_2 - a_1 = b(\log x_2 - \log x_1)$$

If $a_2 - a_1 = 1$, then the change in log parasite intensity at which p percentage of hosts survive is $1/b$.

To estimate the parameters in equation 1, the Adjei Method first calculates the expected number of hosts with a given parasite load x by using the equation $g(x; \mu_p, k_p) * N_p$,

where $g(x; \mu_p, k_p)$ is the negative binomial pre-mortality distribution. Second, the observed and predicted number of hosts with x parasites are paired as a single data point and the method then assumes that this data point follows a binomial distribution with the total number of "trials" equal to the predicted number of hosts and the total number of "successes" equal to the observed number of hosts. In some cases, the observed number of hosts is greater than the expected number of hosts and the Adjei Method alters the data so that the observed is equal to the predicted (Adjei et al., 1986). After this questionable manipulation, the (observed, predicted) pairs are fit to a standard Generalized Linear Model (McCullagh and Nelder, 1989) with a binomial response variable and a logistic link function given by equation 1. This model provides estimates for parameters a, b and LD_{50}.

While not included in the original implementation of the Adjei Method, a χ^2 test with a degrees of freedom of 1 can be used to assess whether a GLM model that includes parasite intensity as a predictor of host survival probability is a "better" model than a GLM without this predictor. This allows the Adjei Method to determine whether PIHM is a significant factor in a host-parasite system.

The Adjei Method's most glaring deficiency is the need to alter the observed data in order to fit the model into the binomial GLM framework. A second more subtle problem with the Adjei Method is the potential need to bin data in order to predict greater than one host in a given parasite intensity class. For example, if the total number of hosts pre-mortality was 50, the mean number of parasites per host pre-mortality was 100 and the aggregation parameter was 1, applying the equation $g(x; \mu_p = 100, k_p = 1) * 50$ would result in less than 1 individual in all parasite intensities x. In other words, the Adjei Method cannot be applied to samples with either very high mean parasite loads, small sample sizes, or both without some sort of binning of the data. While this is not a flaw

per se, it does add a certain level of subjectivity (i.e. which bins should you use?) to a method that already has serious potential issues. In this analysis, we always assume the Adjei Method is not binning the data, though we provide code for applying the binning method in Section 2.C.

2.C Code and unit tests for estimating parasite induced host mortality

Python code, unit tests, and a help file for the Crofton Method, the Adjei Method and the Likelihood Method can be found at `https://github.com/mqwilber/parasite_mortality`

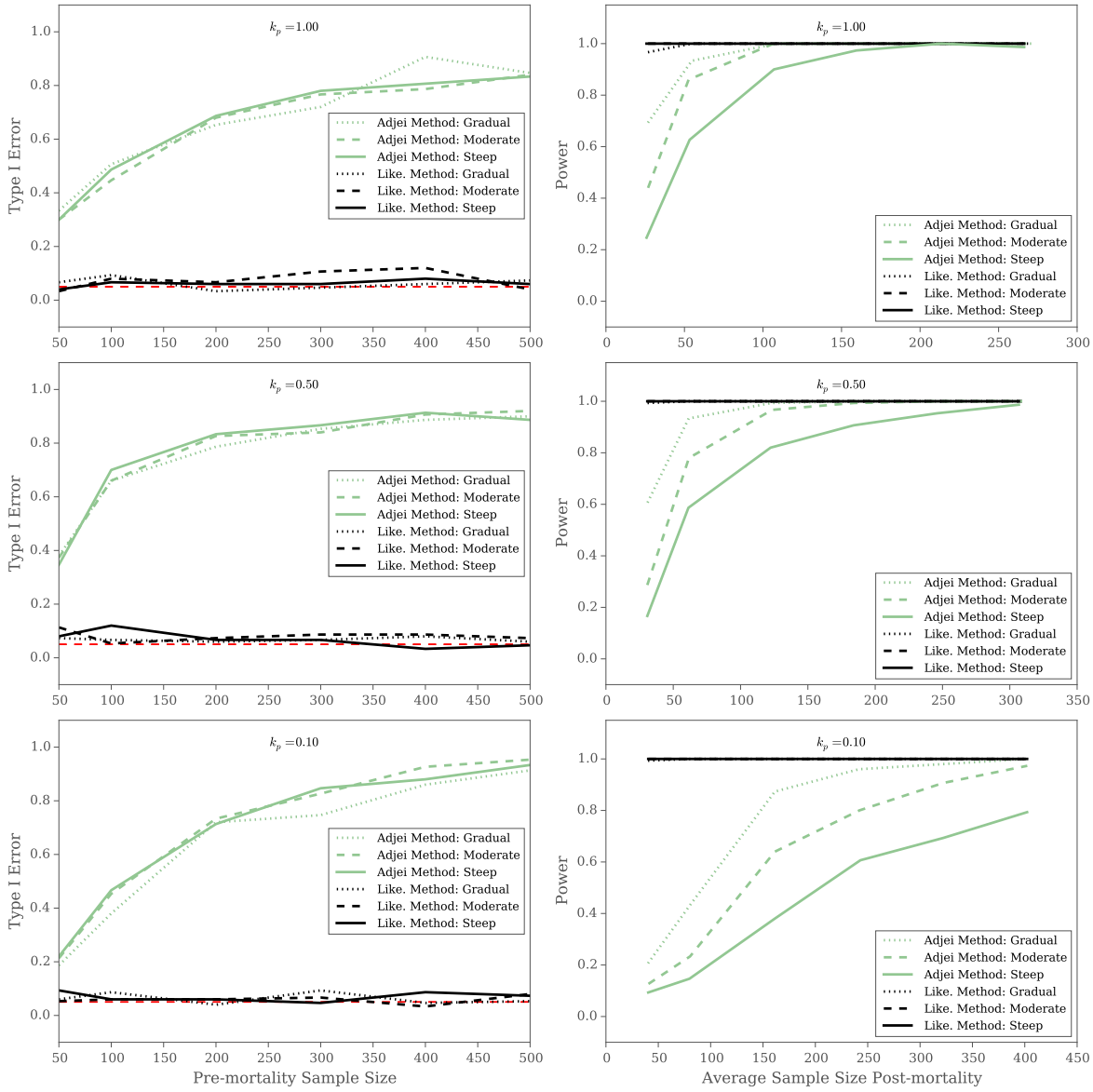


Figure 2.S1: The type I error rate and the power of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 10$ for various shapes of the host survival function and levels of aggregation k_p. The first column gives the type I error rate of each method for falsely detecting PIHM when none is present. The red line gives the the pre-set type I error rate of $\alpha = 0.05$. The second column gives the power of a given method to detect PIHM when it is actually occurring.

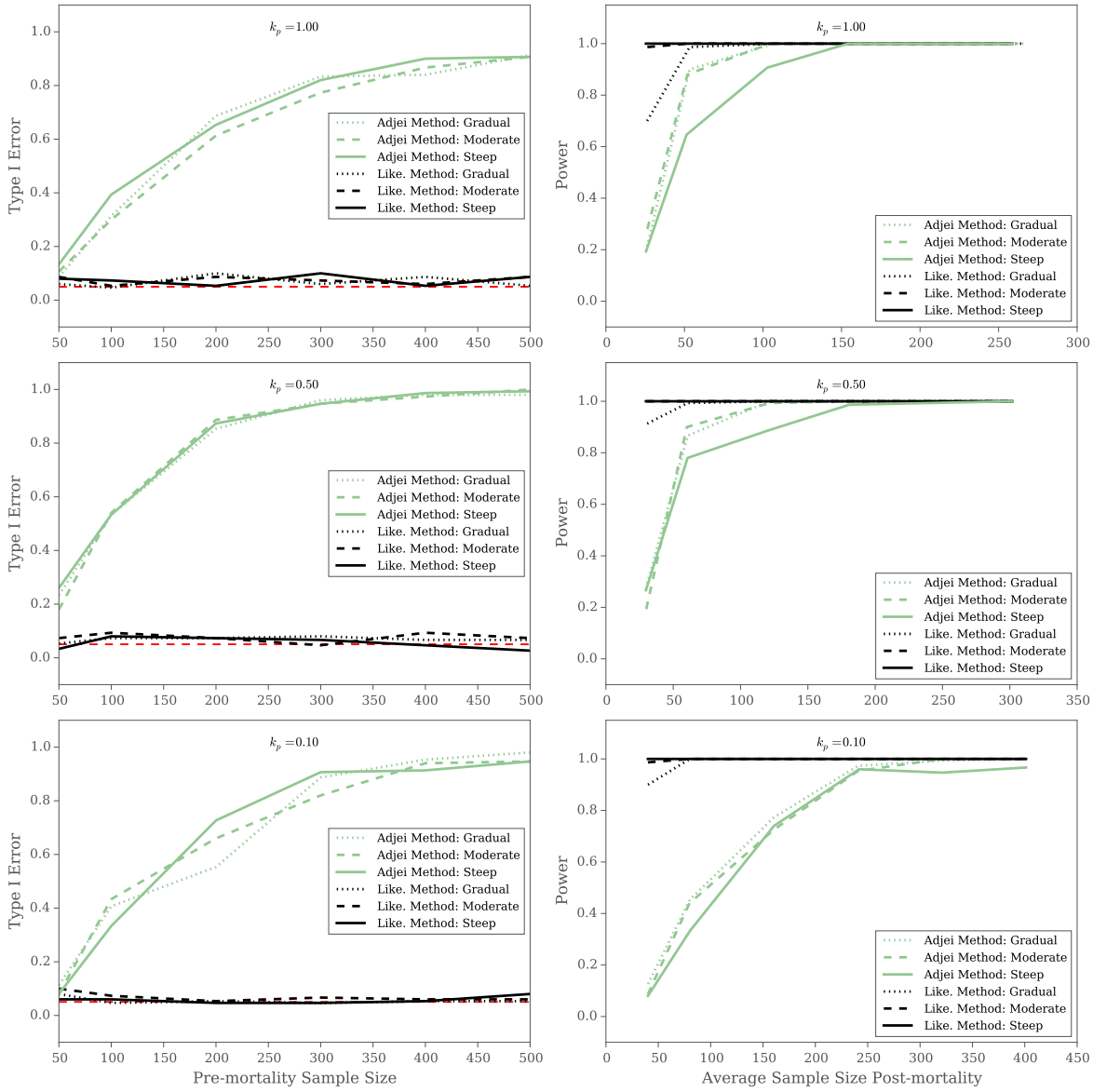


Figure 2.S2: The type I error rate and the power of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 50$ for various shapes of the host survival function and levels of aggregation k_p. The first column gives the type I error rate of each method for falsely detecting PIHM when none is present. The red line gives the the pre-set type I error rate of $\alpha = 0.05$. The second column gives the power of a given method to detect PIHM when it is actually occurring.

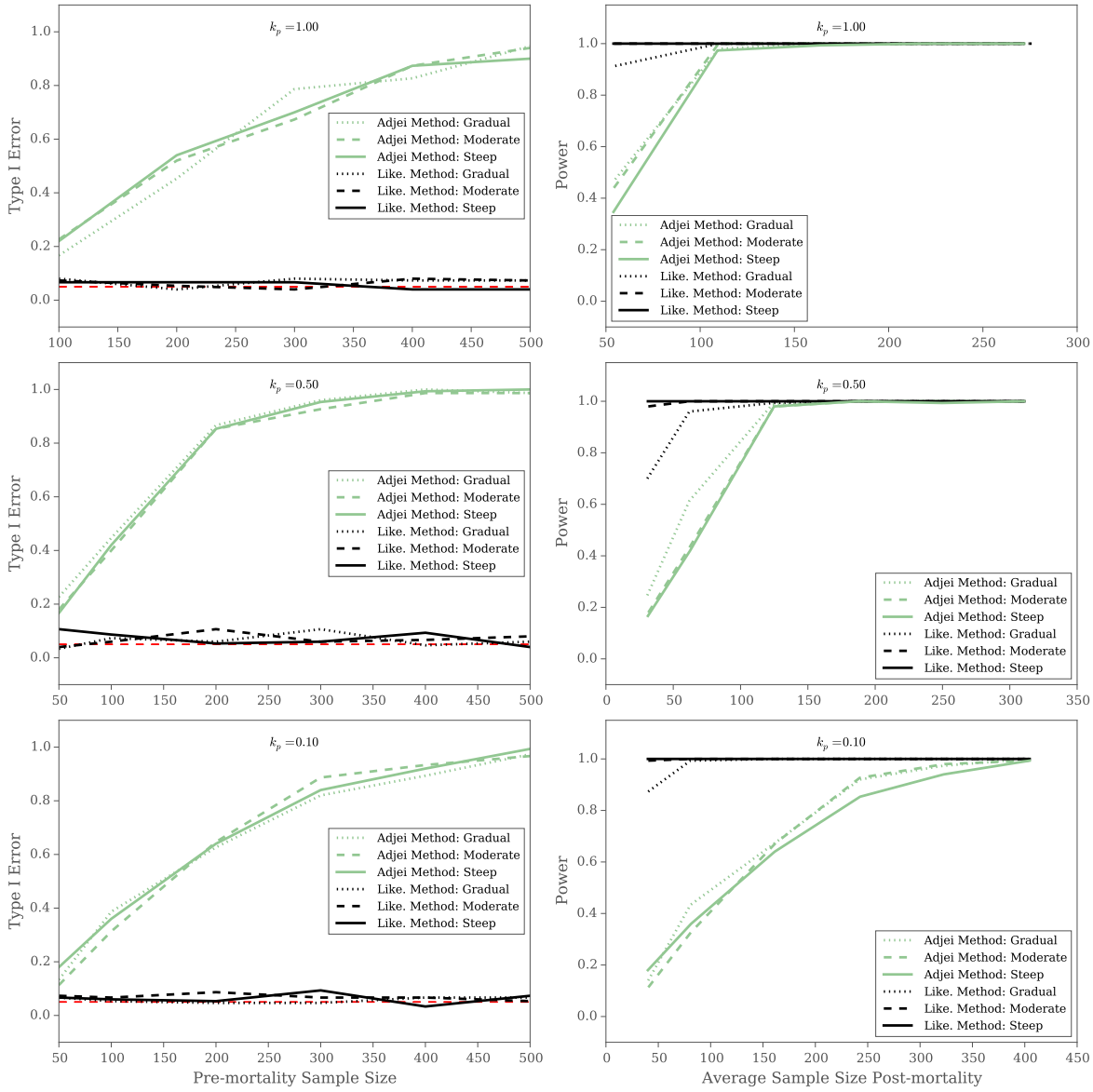


Figure 2.S3: The type I error rate and the power of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 100$ for various shapes of the host survival function and levels of aggregation k_p. The first column gives the type I error rate of each method for falsely detecting PIHM when none is present. The red line gives the the pre-set type I error rate of $\alpha = 0.05$. The second column gives the power of a given method to detect PIHM when it is actually occurring.

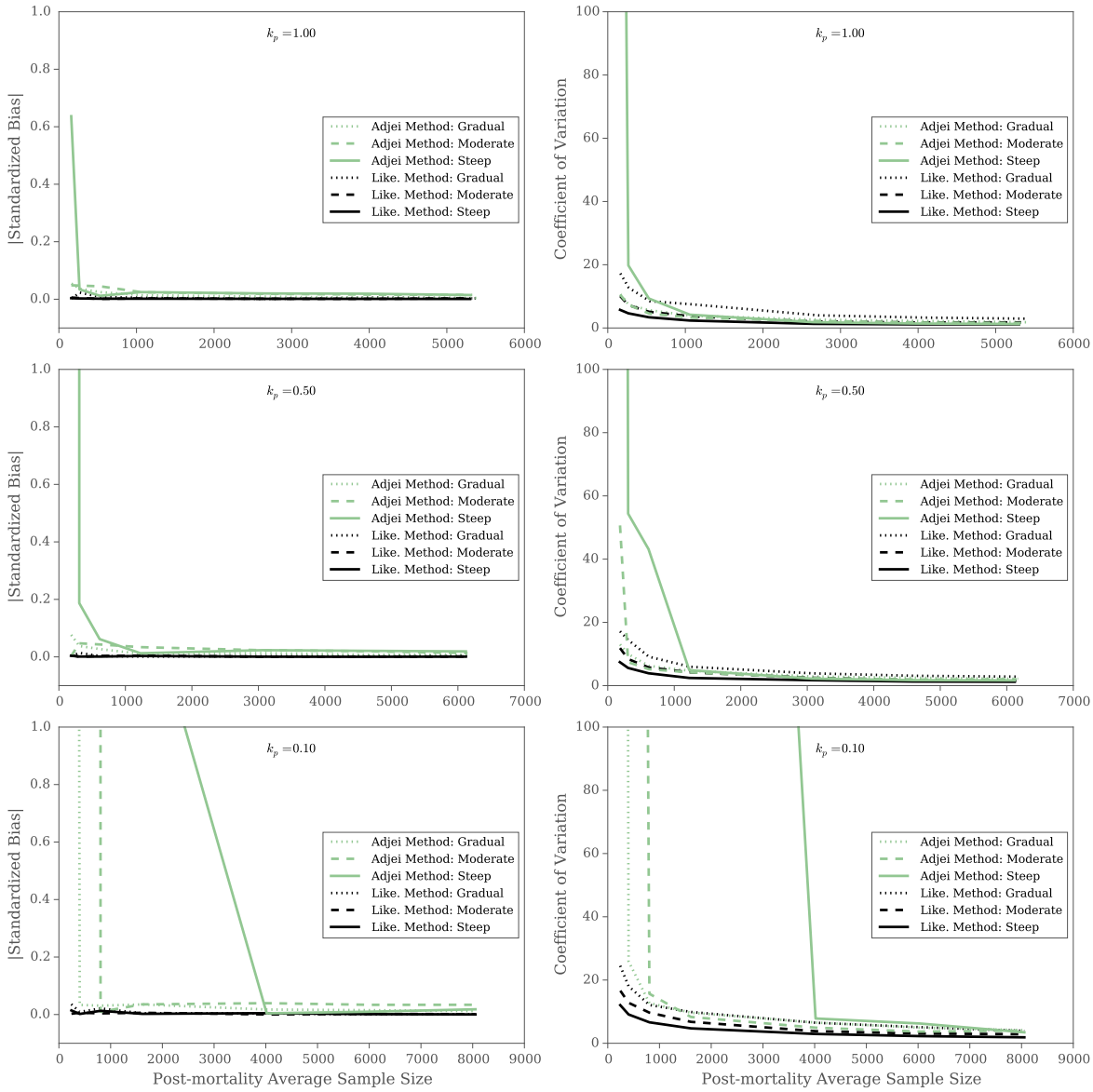


Figure 2.S4: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 10$ for various shapes of the host survival function and levels of aggregation k_p when estimating LD_{50}. The first column gives the bias of each method's LD_{50} estimate over 150 simulations. The second column gives the precision of each method's LD_{50} estimate over 150 simulations.

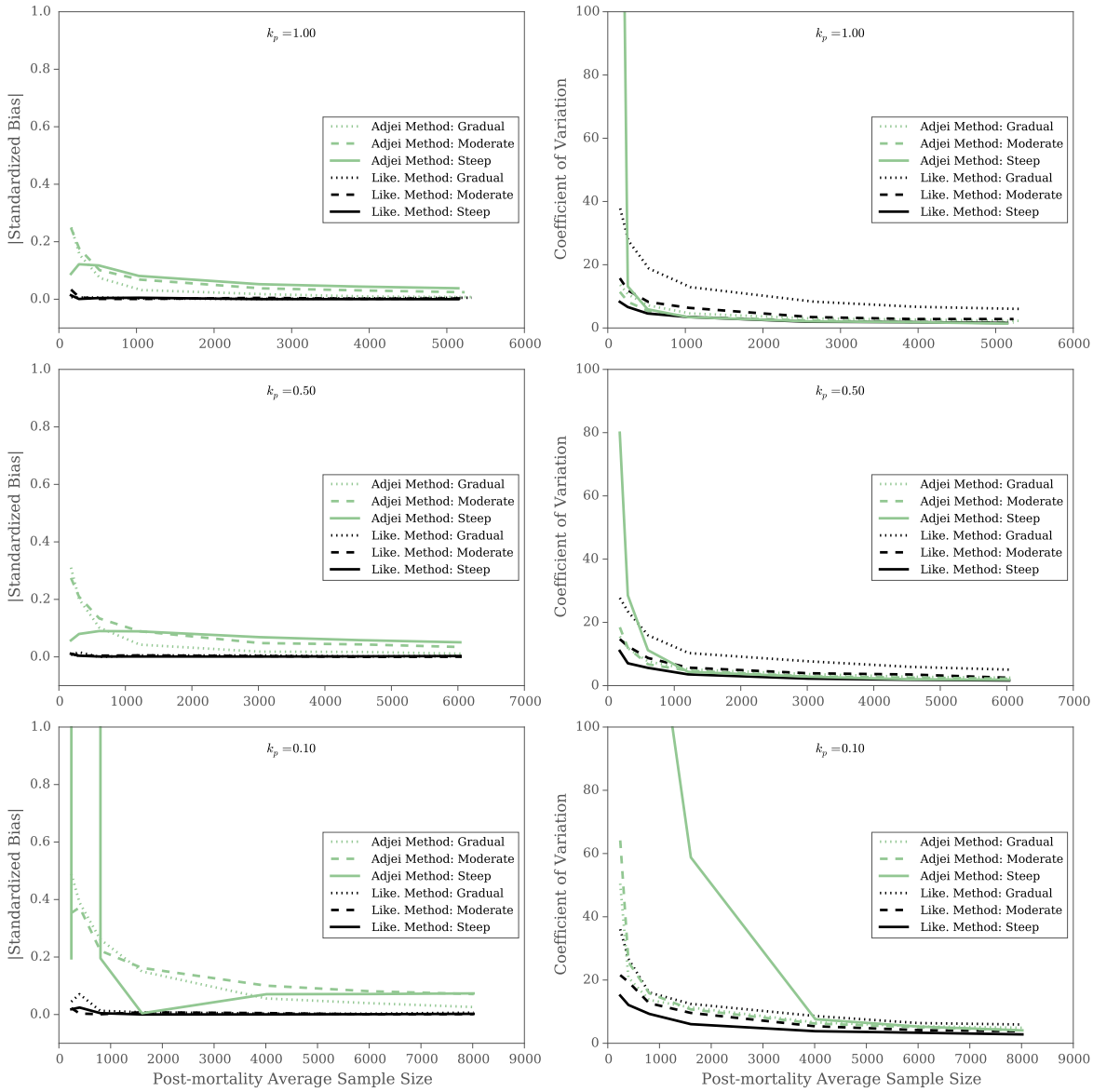


Figure 2.S5: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 50$ for various shapes of the host survival function and levels of aggregation k_p when estimating LD_{50}. The first column gives the bias of each method's LD_{50} estimate over 150 simulations. The second column gives the precision of each method's LD_{50} estimate over 150 simulations.

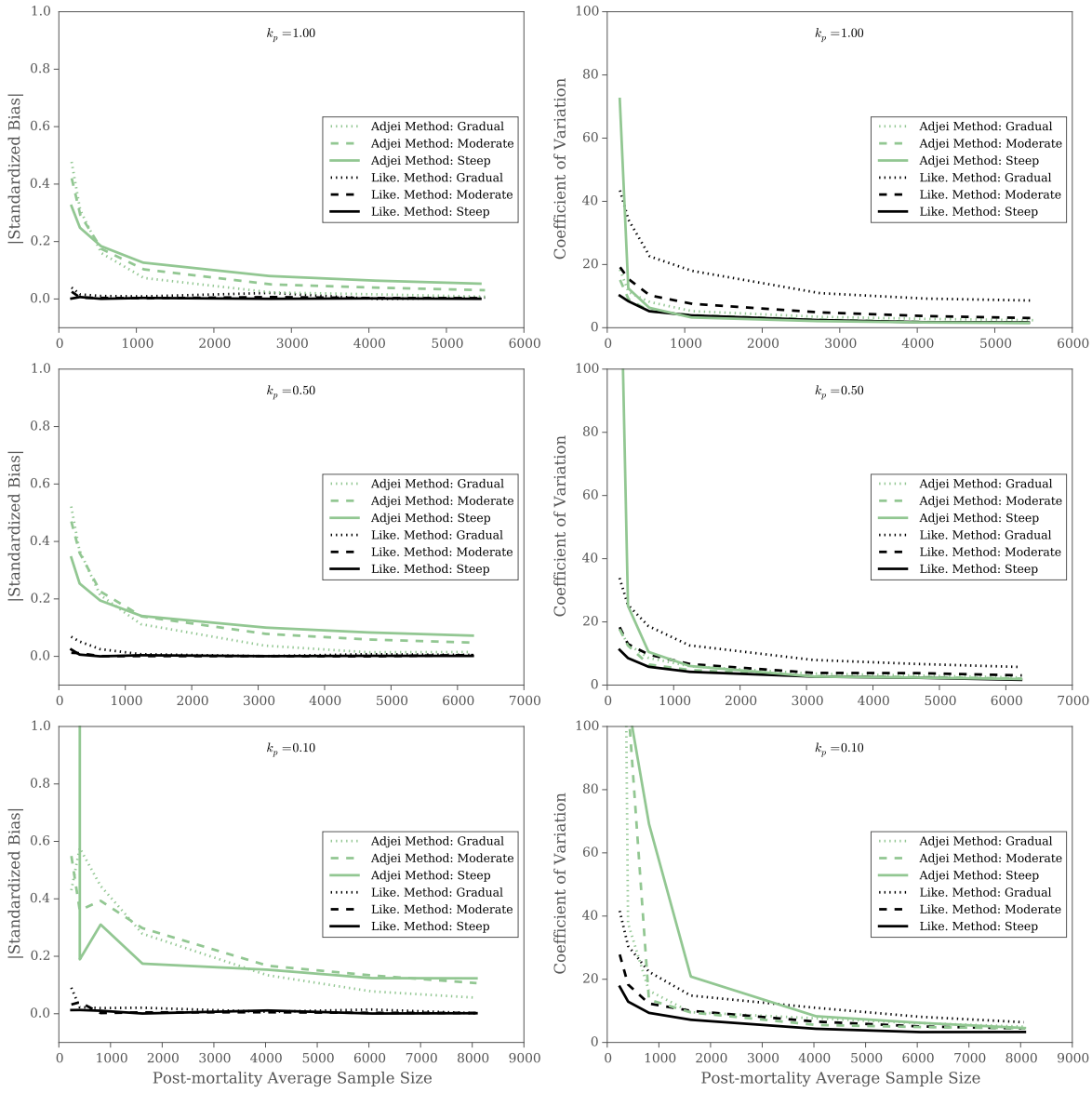


Figure 2.S6: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 100$ for various shapes of the host survival function and levels of aggregation k_p when estimating LD_{50}. The first column gives the bias of each method's LD_{50} estimate over 150 simulations. The second column gives the precision of each method's LD_{50} estimate over 150 simulations.

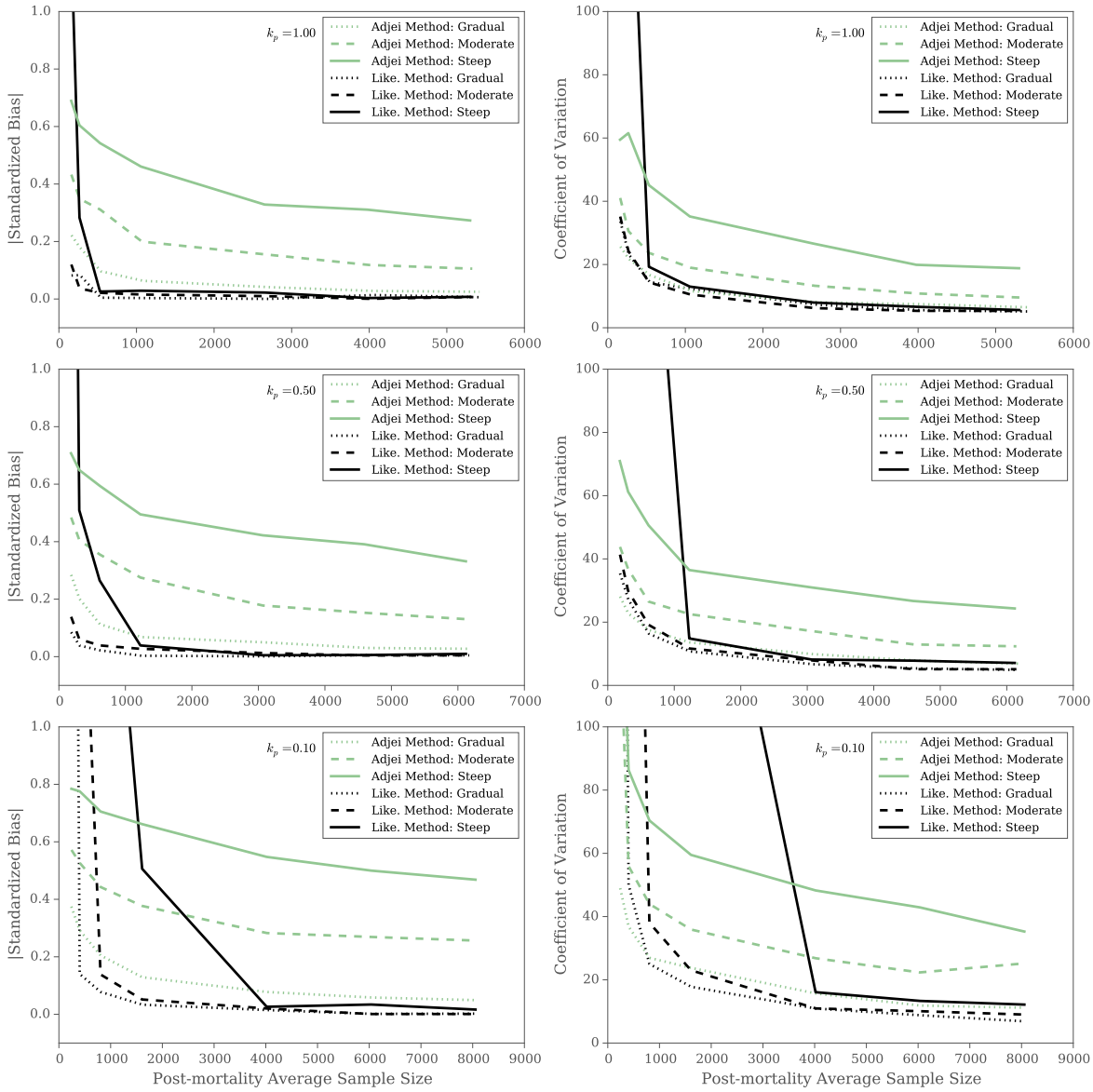


Figure 2.S7: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 10$ for various shapes of the host survival function and levels of aggregation k_p when estimating the a parameter of the host survival function. The first column gives the bias of each method's a estimate over 150 simulations. The second column gives the precision of each method's a estimate over 150 simulations.

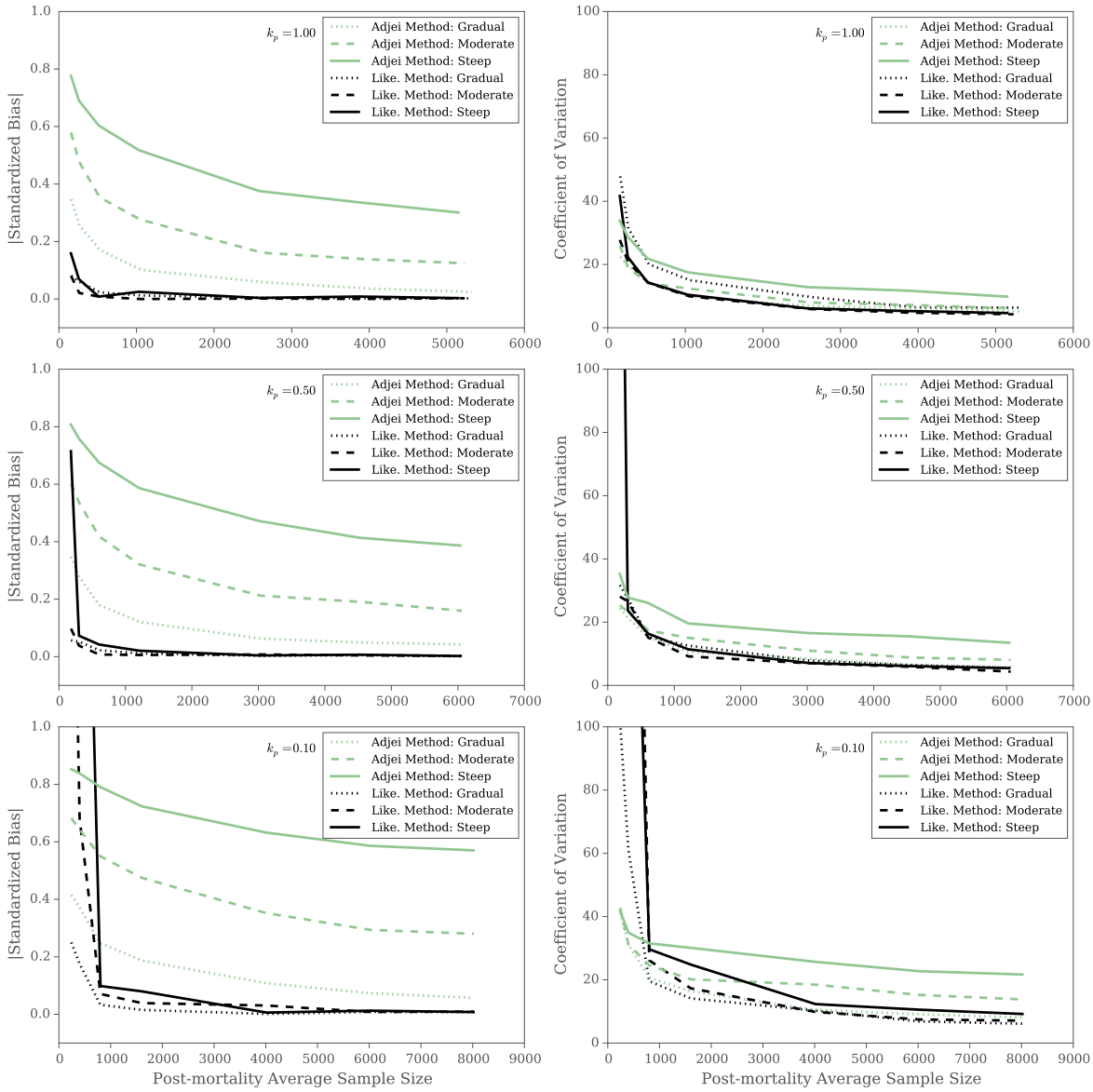


Figure 2.S8: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 50$ for various shapes of the host survival function and levels of aggregation k_p when estimating the a parameter of the host survival function. The first column gives the bias of each method's a estimate over 150 simulations. The second column gives the precision of each method's a estimate over 150 simulations.

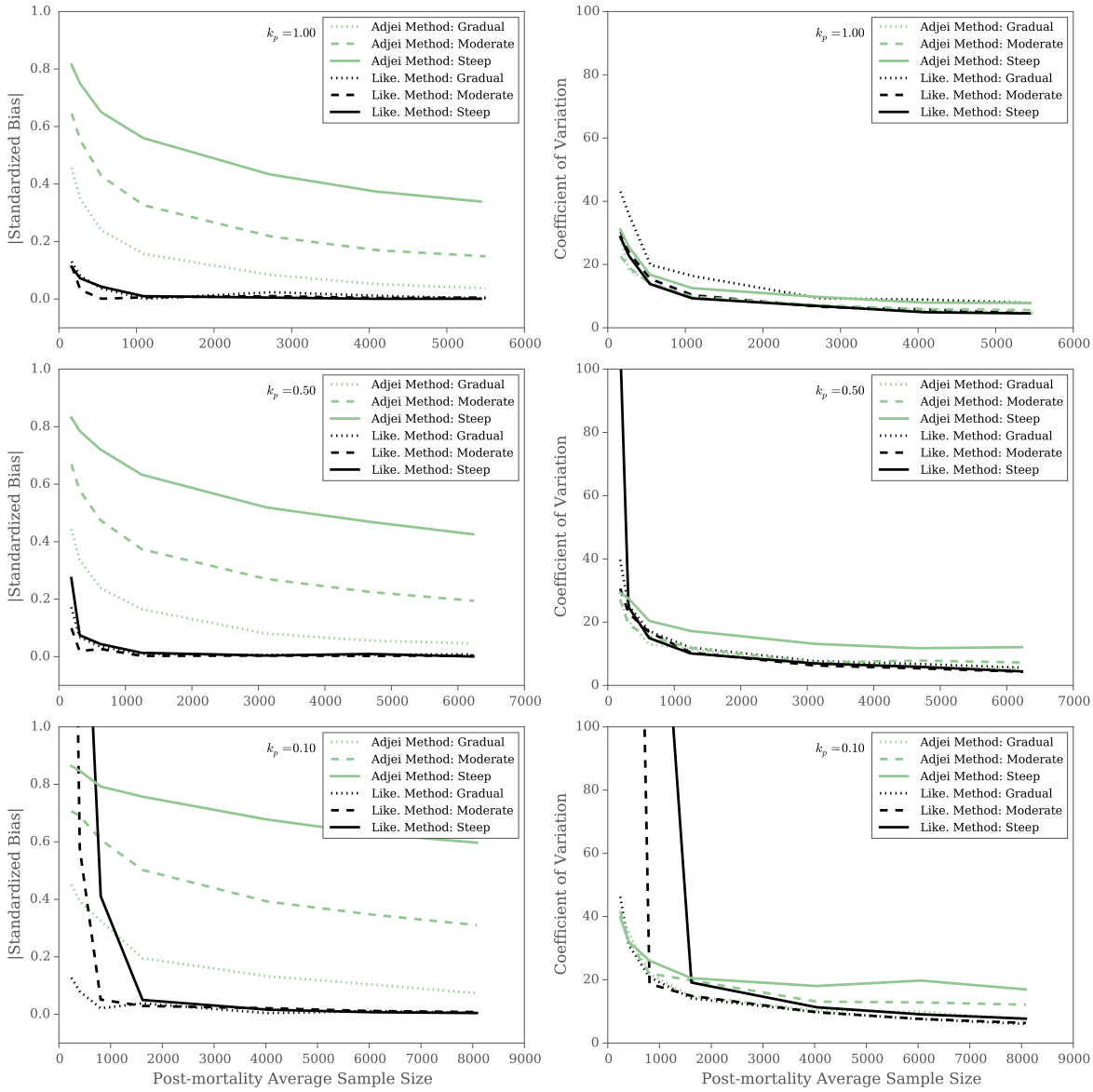


Figure 2.S9: The bias and the precision of the Likelihood Method (black lines) and the Adjei Method (green lines) when $\mu_p = 100$ for various shapes of the host survival function and levels of aggregation k_p when estimating the a parameter of the host survival function. The first column gives the bias of each method's a estimate over 150 simulations. The second column gives the precision of each method's a estimate over 150 simulations.

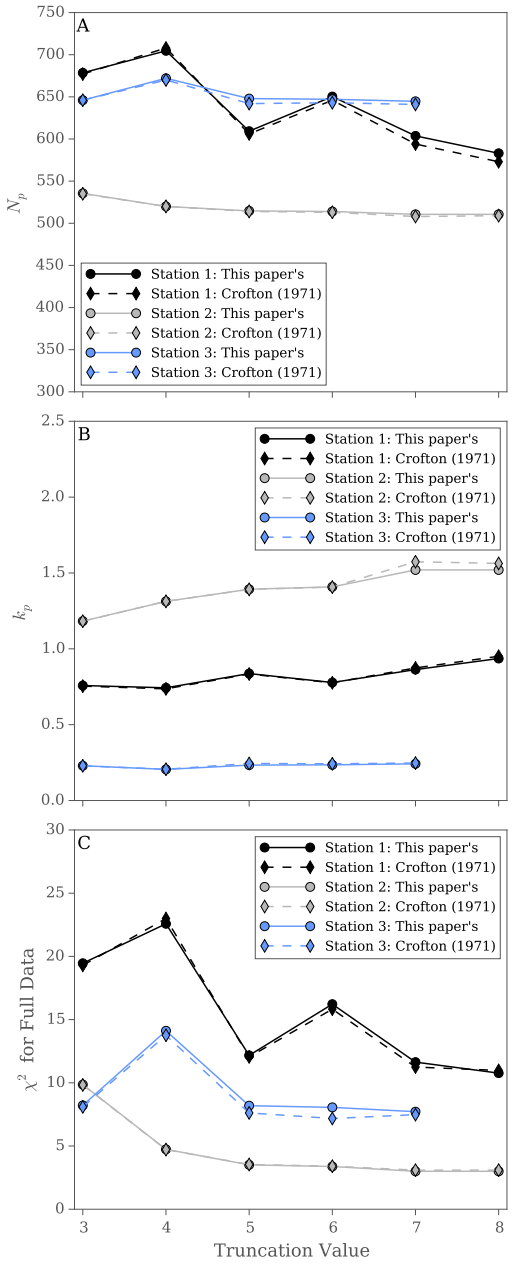


Figure 2.S10: A comparison of this paper's implementation (solid lines, circles) of the Crofton Method with the results given in Crofton (1971) (dashed lines, diamonds). (A) compares the predicted number of hosts in a population pre-mortality (N_p). (B) compares the predicted parasite aggregation pre-mortality (k_p). (C) compares the χ^2 statistic for each implementation. Three of the 6 stations fit by Crofton are shown here and all show that our implementation gives very similar results to those given by Crofton.

Chapter 3

Dynamic parasite aggregation reduces parasite regulation of host populations and the stability of host-parasite interactions

3.1 Abstract

Macroparasites can have significant effects on the dynamics of host populations that include limiting the growth of a host population and altering the stability of a host-parasite equilibrium. An important factor determining the dynamical effects that parasites have on host populations is how aggregated the distribution of parasites is across hosts in the population. While canonical host-macroparasite models assume that parasite aggregation is fixed, empirical and theoretical-work suggests that fixed aggregation is likely the exception rather than the rule. Here I assumed that parasite aggregation varies according to an empirically-supported constraint-based model, in contrast to the canonical assumption of fixed parasite aggregation. I then explored how this assumption alters the ability of parasites to regulate and suppress host populations, as well as stabilize the equilibrium of host-parasite interactions. The model with dynamic aggregation predicts that parasites are less likely to regulate host populations and the host-parasite equilibrium is less likely to be stable compared to the models that assume fixed aggregation. These results suggest that the theoretical ability of parasites to regulate host populations and stabilize the host-parasite equilibrium may be more rare than previously thought, providing an additional explanation for the general lack of empirical evidence for parasites as a primary factor regulating host populations.

3.2 Introduction

Macroparasites can alter the dynamics of wildlife populations (Hudson et al., 1998; Tompkins and Begon, 1999; Albon et al., 2002). One notable way that macroparasites can do this is by suppressing host population abundance (Anderson and May, 1978; Rosà

and Pugliese, 2002; Albon et al., 2002; Tompkins et al., 2002). In simplest terms this means that the presence of a parasite reduces a host population to a lower abundance than would be observed by a similar host population without the parasite. The ability of a parasite to suppress a host population has important implications for managing wildlife populations of conservation and economic interest (Murdoch et al., 1985; Peterson, 2004; Tompkins et al., 2011).

In addition to suppressing the host population, particular macroparasite traits can stabilize or destabilize a host-parasite equilibrium (May and Anderson, 1978; Rosà and Pugliese, 2002; Tompkins et al., 2002). For example, density-dependent interactions of parasites within a host can stabilize the host-parasite equilibrium (Anderson and May, 1978), whereas parasite-induced reductions in host fecundity can destabilize the equilibrium (May and Anderson, 1978; Diekmann O. and Kretzschmar, 1991). While these different dynamical effects have been shown theoretically, empirical examples of parasites directly affecting host dynamics remain limited to a few high-profile studies (Scott, 1988; Hudson et al., 1998; Albon et al., 2002; Redpath et al., 2006). In part, this is due to the logistical difficulties of manipulative experiments that are needed to definitively link parasites to changes in host population dynamics (Tompkins et al., 2002, 2011). Because of this difficulty, epidemiological models are critical for predicting the conditions under which parasites may be expected to suppress and stabilize host-parasite systems (Dobson and Hudson, 1992; Albon et al., 2002; Tompkins et al., 2002).

In addition, epidemiological models can provide insight into the conditions under which a parasite can regulate a host population. Following previous work in host-macroparasite systems, I specifically define regulation as the ability of a parasite to prevent a host population from growing without bound (Anderson and May, 1978; Rosà and Pugliese, 2002). While in reality other ecological factors will limit host population

size if parasites fail to do so, this simple definition of regulation is often used in host-macroparasite models (Diekmann O. and Kretzschmar, 1991; Albon et al., 2002; Rosà and Pugliese, 2002). In this study I distinguish regulation from suppression, which refers to the ability of a parasite to reduce equilibrium host abundance, conditional on regulation. Particular attributes of macroparasites affect their ability to regulate a host population in models. For example, increasing parasite reproductive rate increases the ability of parasites to regulate host populations, while increasing parasite virulence decreases the ability of parasites to regulate host populations (Anderson and May, 1978).

Parasite aggregation is an attribute of host-macroparasite systems that has implications for the potential of a parasite to regulate a host population (i.e. prevent unbounded growth), suppress a host population (i.e. reduce the equilibrium host abundance) and stabilize the host-parasite equilibrium (Anderson and May, 1978; Kretzschmar and Adler, 1993). Aggregation is the nearly ubiquitous pattern in parasite ecology that many hosts have few parasites and few hosts have many parasites (Shaw and Dobson, 1995; Shaw et al., 1998). Specifically, this means that host-parasite distributions often have a variance to mean ratio greater than one and are highly right-skewed. In host-parasite models, increasing aggregation decreases a parasite's ability to regulate a host population, but, if regulation is possible, then increasing aggregation increases the stability of the host-parasite equilibrium, but decreases the suppression of the host population by the parasite (Anderson and May, 1978). The effect of aggregation on regulation, suppression, and stability also interacts with other processes operating in the host-parasite system. For example, highly aggregated and highly virulent parasites are unable to regulate a host population as many parasites are removed from the population upon the death of a host, which occurs frequently (Anderson and May, 1978). On the other hand, if parasites reduce host fecundity, then higher levels of aggregation are

needed to stabilize the host-parasite equilibrium (Tompkins et al., 2002). Thus aggregation plays an important role in host-parasite dynamics both independently and through its interaction with other host- and parasite-related vital rates.

Many mechanisms can affect parasite aggregation, including clumped infections (Isham, 1995), parasite-induced mortality (Barbour and Pugliese, 2000), host heterogeneity (Wilson et al., 2002; Gourbière et al., 2015), and the balance between parasite immigration rate and birth rate (Fowler and Hollingsworth, 2016). Due to multiple interacting mechanisms shaping patterns of parasite aggregation, one might assume that predicting aggregation would require highly system-specific models. Surprisingly, this is not the case. Recent work has shown that because many mechanisms interact to affect aggregation, general statistical rules can predict observed patterns of aggregation across different host-parasite systems (Wilber et al., 2017). For example, simple constraint-based models developed in community ecology can predict patterns of aggregation across a range of amphibian (host)-trematode (macroparasite) systems (Wilber et al., 2017). Constraint-based models are rooted in the concept of maximum entropy and predict the most likely distribution given a set of constraints (Haegeman and Etienne, 2010; Harte, 2011; Locey and White, 2013; Harte and Newman, 2014). In host-parasite systems, there are two inherent constraints on any observed parasite distribution: the total number of parasites P and the total number of hosts H in the distribution (Wilber et al., 2017). This means that the only possible distribution of P parasites amongst H hosts that could be observed is one that is consistent with these constraints. The most-likely distribution given these constraints is the one that can be realized in the largest number of ways (Jaynes, 1982).

An important prediction from constraint-based models is that the level of parasite aggregation changes as the total number of hosts H and parasites P change (Locey

and McGlinn, 2013; Johnson and Wilber, 2017). This is important because theoretical results have shown that when aggregation varies with H and P, the conditions for a stable equilibrium in host-parasite systems are altered, compared to the canonical assumption of fixed aggregation (Adler and Krestzschmar, 1992; Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002). For example, Kretzschmar and Adler (1993) showed that aggregation itself is not sufficient to stabilize a host-parasite system. Rather, aggregation needs to be an increasing function of mean parasite load (P/H) to stabilize the host-parasite equilibrium. Despite this well-known importance of aggregation that depends on P and H, there is no theory that predicts the precise shape of this relationship. Constraint-based theory provides such a prediction.

In this study I explore the effects of the constraint-based aggregation on the dynamics of host-parasite systems. In particular, this study asks two questions. First, how do constraint-based models for parasite aggregation affect the ability of a parasite to regulate and suppress a host population? Second, how does constraint-based aggregation affect the stability of the host-parasite equilibrium? This study shows that accounting for empirically-supported constraint-based models of aggregation reduces the parameter space in which parasites are predicted to regulate host populations and stabilize the host-parasite equilibrium, compared to the standard assumption of fixed aggregation.

3.3 Models

3.3.1 Overview

The goal of this study was to understand whether constraint-based aggregation changed how aggregation affected host-parasite dynamics, compared to the canoni-

cal assumption of fixed parasite aggregation. To this end, I extended a simple host-macroparasite model developed in previous studies (described below, Anderson and May, 1978; Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002). This simple model excludes important processes affecting host-macroparasite systems such as host heterogeneity (beyond that captured by heterogeneity in parasite intensity), clumped infections, adaptive immunity, and logistic host growth, all of which have been explored previously (Isham, 1995; Pugliese and Rosa, 1995; Pugliese et al., 1998; Rosà and Pugliese, 2002). These processes were excluded to allow for direct comparison between the dynamics of host-macroparasite systems following constraint-based aggregation and those predicted by the canonical host-macroparasite model with fixed aggregation (Anderson and May, 1978; Rosà and Pugliese, 2002). By first analyzing the behavior of the simple model in the absence of the mechanisms mentioned above, the impact of constraint-based aggregation on host-parasite dynamics can be understood without the confounding effects of other stabilizing and destabilizing mechanisms.

In the following sections I describe two host-macroparasite models, one with fixed aggregation and one with aggregation following constraint-based predictions. I then compare and contrast predictions from the two models regarding the ability of a parasite to regulate and suppress the host population and stabilize the host-parasite equilibrium.

3.3.2 The host-macroparasite model

Consider the following host-macroparasite model where H is the abundance of hosts in a population and P is the abundance of directly reproducing macroparasites across all hosts in the population (Anderson and May, 1978; Diekmann O. and Kretzschmar, 1991; Rosà and Pugliese, 2002)

$$\frac{dH}{dt} = -dH - \alpha P + bH \sum_{i=0}^{\infty} r_i (1-\xi)^i \tag{3.1}$$

$$\frac{dP}{dt} = -(d+\mu)P + \phi H - \alpha H \sum_{i=0}^{\infty} i^2 r_i \tag{3.2}$$

b is host birth rate, d is host death rate, α is the rate of parasite-induced host mortality, μ is parasite death rate, and ϕ is the rate at which parasites are acquired from the environment. ϕ takes the functional form $\frac{\lambda P}{H_0+H}$, which is a result of assuming that dynamics of the free-living stage of the parasite occur on a much faster time scale than the within-host dynamics (Anderson and May, 1978; Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002). $\sum_{i=0}^{\infty} i^2 r_i = E[i^2]$ is the second moment of the host-parasite distribution where r_i is the probability of a host having i parasites. This term is a result of the assumption that parasite-induced death rate increases linearly with the number of parasites per host. $\sum_{i=0}^{\infty} r_i (1-\xi)^i$ is the probability generating function of the host-parasite distribution where $0 \leq (1-\xi) \leq 1$ is the multiplicative reduction of host fecundity due to the parasite ($\xi = 0$ is no reduction). This two dimensional system of equations is an approximation of an infinite series of ordinary differential equations (ODEs) that tracks the number of hosts with $i = 0, 1, \ldots, \infty$ parasites (Anderson and May, 1978; Kretzschmar and Adler, 1993). The relationship between equations 1 and 2 and the infinite series of ODEs are discussed at length elsewhere (Rosà and Pugliese, 2002; Cornell, 2010).

This model makes a number of assumptions, two of which deserve immediate attention. First, the model assumes that there is no density-dependence in host birth or death rates. This means that host population size can only be regulated by parasites and, upon escaping parasite control, hosts will increase exponentially (for $b > d$, Anderson and May, 1978). This is a useful simplification for precisely defining para-

site regulation – preventing a host population from increasing exponentially (Rosà and Pugliese, 2002). Second, the model makes the assumption that the host population is homogeneous. In reality host populations have stage- and age-structure and hosts vary in immunity and behavior, all of which have implications on host-parasite dynamics and the patterns of parasite aggregation (Pacala and Dobson, 1988; Isham, 1995; Chan and Isham, 1998; Wilson et al., 2002; Calabrese et al., 2011; Johnson and Hoverman, 2014). However, because the goal is to make general conclusions about how constraint-based models of aggregation can affect host-parasite dynamics, I follow the example of similar theoretical studies and ignore host heterogeneity.

3.3.3 Modeling parasite aggregation

To close the system of equations given above, it is often assumed that the parasite distribution follows a negative binomial distribution – a flexible two parameter distribution that fits observed, aggregated host-parasite distributions (Anderson and May, 1978; Shaw et al., 1998). The second moment of a negative binomial distribution is given by $E[i^2] = \frac{P}{H} + \frac{P^2}{H^2}\frac{k+1}{k}$, where k is the aggregation parameter of the negative binomial distribution and decreasing k indicates increasing aggregation ($k \in (0, \infty)$). The probability generating function for a negative binomial distribution is given by $G(1-\xi) = \sum_{i=0}^{\infty} r_i(1-\xi)^i = \left(\frac{kH}{\xi P + kH}\right)^k$.

Plugging these two equations into 1 and 2 leads to the host-macroparasite model with fixed k (henceforth the Fixed k Model; Diekmann O. and Kretzschmar, 1991; Rosà and Pugliese, 2002)

$$\frac{dH}{dt} = -dH - \alpha P + bH\left(\frac{kH}{\xi P + kH}\right)^{k} \tag{3.3}$$

$$\frac{dP}{dt} = -(d+\mu)P + \phi H - \alpha H\left(\frac{P}{H} + \frac{P^2}{H^2}\frac{k+1}{k}\right) \tag{3.4}$$

While assuming a fixed k is a common simplifying assumption (e.g. Dobson and Hudson, 1992; Townsend et al., 2009), it is well-known that parasite aggregation can vary over time (Scott, 1987; Boag et al., 2001). Deterministic host-parasite models have captured the dynamic nature of aggregation using two different approaches. The first approach uses a three dimensional approximation of the infinite series of differential equations such that aggregation (typically the variance to mean ratio) is explicitly modeled as a dynamic state variable (Adler and Krestzschmar, 1992; Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002). While useful, this approach does not obviate the need to assume some distribution (often a negative binomial) to close the three-dimensional system of equations (Rosà and Pugliese, 2002; Cornell, 2010).

Another approach is to use the two dimensional approximation given by equations 1 and 2, assume a negative binomial distribution, and then allow aggregation (given by k) to be a generic function of the total number of parasites and the total number of hosts in a population (Kretzschmar and Adler, 1993). The advantage to this approach is flexibility – aggregation can be driven by both the dynamics of the system affecting H and P and the assumed relationship between k, H and P. However, this flexibility is also a disadvantage as the assumed relationship between k, P, and H is phenomenological (e.g. Kretzschmar and Adler, 1993). Because constraint-based models provide an empirically- and theoretically-supported relationship for how aggregation varies with P and H (Wilber et al., 2017), they can be used to explore how the varying nature of aggregation affects the dynamics of a host-parasite system.

3.3.4 Incorporating constraint-based aggregation

Here I consider the constraint-based model known as "the partition model" for aggregation (Xiao et al., 2015; Wilber et al., 2017). The partition model predicts the most likely host-parasite distribution given the constraints of total host abundance H and total parasite abundance P, as well as unlabeled hosts and unlabeled parasites (see Haegeman and Etienne, 2010; Wilber et al., 2017, for a complete description). In short, this means that, given P parasites and H hosts, there are only a finite number of ways that these parasites can be distributed among the hosts. One realized distribution of parasites among hosts is a configuration. Assuming unlabeled hosts and parasites means that each possible configuration is equally likely. While there is no known analytical solution for the partition model, it can easily be estimated via simulation (Locey and McGlinn, 2013). Fig 3.1A. shows five predicted host-parasite distributions from a partition model with different numbers of P parasites and H hosts. Perhaps not surprisingly, fitting a negative binomial distribution to the expected prediction from a partition model reveals that the general shape of the partition model can be approximately captured by a negative binomial distribution where k varies as a function of P and H (Fig. 3.1A).

Assuming that the partition model can be approximated as a negative binomial distribution where k is a function of P and H, I computed the $k(P, H)$ surface using the following steps. First, I drew 1000 random partitions (i.e. potential configurations of P parasites among H hosts) from all the combinations of partition models with $P = 3, \ldots, 500$ and $H = 3, \ldots, 500$ (248,004 combinations). For any given (P, H) pair, this resulting draw of partitions resulted in matrix with 1000 rows (i.e. 1000 random partitions) and H columns. The probability distribution of a given host having x parasites under the partition model with P and H is $p(x|P, H) = \sum_{m \in F} p(x|m, P, H)p(m|P, H)$

where m is a single partition in the feasible set F. Given the 1000 samples, I approximated $p(x|P, H)$ with a negative binomial distribution by first flattening the 1000 x H matrix into a vector and then estimating the aggregation parameter k of the negative binomial distribution from this $n = 1000H$ "sample" using a corrected moment estimate (Gregory and Woolhouse, 1993).

The above steps provided $\hat{k}$ values for all discrete combinations of P and H on a 498 x 498 grid (P and H are both between 3 - 500). However, to use this surface with equations 1 and 2, $\hat{k}$ needed to be defined when P and H were not integers. To do this, I increased the resolution of the 498 x 498 grid such that P and H both had 10,000 equally spaced points between 3 and 500. I used cubic polynomial interpolation to calculate $\hat{k}$ for all 1 x 10^8 (P, H) on the high resolution grid (Jones et al., 2013). On this high resolution grid, $\hat{k}$ was less than two for 99.9% of the values and generally only greater than two when H and P where less than or equal to 4. This is consistent the vast majority of host-parasite systems where $k < 2$ (Shaw and Dobson, 1995; Shaw et al., 1998). After increasing the resolution, I removed the simulation noise in $\hat{k}(P, H)$ by applying a Gaussian filter with $\sigma = 75$ (Jones et al., 2013). The choice of σ led to a smooth surface and ensured that all $\hat{k} < 2$. While the choice of σ affects the quantitative conclusions presented below (e.g. the exact location of the boundaries in parameter space), the qualitative results remain consistent across all values of σ that I explored. Finally, I assumed that any points beyond the boundary of the surface had the same value as the nearest point on the boundary. This assumption, along with $\sigma = 75$, led to $\hat{k} \approx 2$ for all values of $P < 3$ and $H < 3$, which is equivalent to a fixed k assumption for these values of P and H.

Fig. 3.1B shows the predicted $\hat{k}(P, H)$ surface after increasing the resolution and applying the Gaussian filter. The surface shows that the partition model predicts that k decreases (aggregation increases) with increasing H and k increases (aggregation de-

creases) with increasing P, unless P and H are small in which case k briefly decreases with increasing P. Using this predicted $\hat{k}(P,H)$ surface, in combination with the second moment of the negative binomial distribution given above, the second moment of the partition model can be approximated as

$$E[i^2] \approx (P/H) + \frac{P^2}{H^2}\frac{\hat{k}(P,H)+1}{\hat{k}(P,H)} \tag{3.5}$$

and the generating function of the partition model as

$$G(1-\xi) \approx \left(\frac{H\hat{k}(P,H)}{\xi P + H\hat{k}(P,H)}\right)^{\hat{k}(P,H)} \tag{3.6}$$

Plugging equation 5 and 6 back into equations 1-2 the host-macroparasite model with aggregation following the partition model (henceforth the Feasible k Model) is given by

$$\frac{dH}{dt} = -dH - \alpha P + bH\left(\frac{H\hat{k}(P,H)}{\xi P + H\hat{k}(P,H)}\right)^{\hat{k}(P,H)} \tag{3.7}$$

$$\frac{dP}{dt} = -(d+\mu)P + \phi H - \alpha H\left(\frac{P}{H} + \frac{P^2}{H^2}\frac{\hat{k}(P,H)+1}{\hat{k}(P,H)}\right) \tag{3.8}$$

To answer the questions regarding how constraint-based aggregation affects the host-parasite dynamics, I compared the ability of parasites to regulate and suppress a host population as well as the stability of host-parasite equilibrium under both the Fixed k Model and the Feasible k Model. All analyses were performed in Python v. 2.7.12 and the code to replicate the analyses is provided at

`https://github.com/mqwilber/parasite_regulation`.

3.4 Question 1: How do parasites regulate and suppress a host population?

To facilitate comparison with previous results, I first assumed that parasites had no effect on host fecundity ($\xi = 0$) and only affected host mortality ($\alpha > 0$). I relaxed this assumption in the following section. Under this assumption, the Fixed k Model predicts that parasites are able to regulate a host population when $\alpha \neq 0$ and $\lambda > b+\mu+\alpha+\frac{(b-d)}{k}$ (Anderson and May, 1978). As aggregation increases (k decreases) it becomes more difficult for parasites to regulate the host population (Fig. 3.2).

To compare this result to the ability of the Feasible k Model to regulate the host population, I simulated equations 7 and 8 using Euler's method with $\Delta t = 0.01$ and updated P, H and $\hat{k}(P, H)$ for each Δt time step. I ran the model until either the host and parasite populations reached a finite attractor (regulation) or the host population increased above some large upper limit (no regulation).

The ability of a parasite to regulate a host population under the Feasible k Model was reduced compared the Fixed k Model (Fig. 3.2). However, when parasite reproductive rate λ was low and parasites were highly aggregated under the Fixed k Model (e.g. $k = 0.2$), there were regions of the parameter space in which the parasites in the Fixed k Model did not regulate the host population, but those in the Feasible k Model did. This was only a small portion of the parameter space and as the reproductive rate of the parasite (λ) increased, the Feasible k Model again showed a reduced region in which parasites could regulate the host population (Fig. 3.2A-C).

These differences between the Fixed k Model and the Feasible k Model can be understood in terms of a positive feedback loop between aggregation and the total abundance of hosts. Under the Fixed k Model, an increase in α increases the equilibrium host

abundance H^* (Fig. 3.3A), but has no corresponding affect on aggregation (Fig. 3.3C). Under the Feasible k Model an increase in α also leads to an increase in H^* (Fig. 3.3A), which generally leads to a corresponding decrease in k (Fig. 3.3C). This increase in aggregation reduces the ability of the parasite to suppress the host population which leads to an increase in H^*, which in turn leads to a decrease in k. This eventually leads to a complete inability of parasites to regulate the host population as small increases in H lead to large decreases in k when H is relatively small (e.g. $H < 100$, Fig. 3.1B, Fig. 3.3C).

As noted above and shown in Fig. 3.3, the ability of the parasite to suppress equilibrium host abundance also varies between the Fixed k Model and the Feasible k Model. For low values of parasite pathogenicity, the Feasible k Model predicts increased host suppression (i.e. reduced equilibrium host abundance) compared to a Fixed k Model with $k = 1$ (Fig. 3.3A). This is because the Feasible k Model predicts $k > 1$ for these low levels of parasite pathogenicity (Fig. 3.3C), which leads to increased suppression of the host population relative to $k = 1$. However, as pathogenicity increases the Feasible k Model predicts that k generally decreases (Fig. 3.3C). Eventually $k < 1$ and the Feasible k Model shows less suppression of host abundance than the Fixed k Model with $k = 1$. This lack of suppression can be further understood by examining how parasite efficiency, defined as the negative log percentage of unparasitized hosts divided by equilibrium parasite abundance (Singh et al., 2009), changes between the Fixed k and Feasible k Model. Fig. 3.3D shows that while parasites are initially more efficient under the Feasible k Model than the Fixed k Model with $k = 1$, their level of efficiency decreases faster under the Feasible k Model with increasing pathogenicity. This decreasing efficiency with increasing pathogenicity leads to reduced host suppression under both models, but the faster decrease in efficiency under the Feasible k Model leads to parasites

failing to regulate host populations at lower levels of parasite pathogenicity (α).

3.5 Question 2: How does dynamic aggregation affect the stability of the host-parasite equilibrium?

3.5.1 Stability properties without a parasite reduction in host fecundity ($\xi = 0$)

To understand the nature of the dynamics in the regulated region of Figure 3.2, I performed a local stability analysis on both the Fixed k Model and the Feasible k Model. I did this by numerically calculating the equilibrium of a given model under a particular set of parameters and then examining characteristics of the Jacobian matrix at that equilibrium (Nisbet and Gurney, 1982). The results from the Fixed k Model are well-known – when a parasite with aggregation $0 < k < \infty$ is able to regulate a host population, the resulting system has a locally stable equilibrium (Fig. 3.2, Anderson and May, 1978). The Feasible k Model showed similar behavior – when the host-macroparasite system was regulated the resulting host-parasite equilibrium was locally stable for the entire parameter space considered (Fig. 3.2). Because the parameter space in which the parasite can regulate the host population is reduced under the Feasible k Model compared to the Fixed k Model (Fig. 3.2), the parameter space in which the host-parasite equilibrium is stable is also reduced under the Feasible k Model.

The stability of the Feasible k Model is consistent with the observation that aggregation is a stabilizing mechanism in host-macroparasite systems (Anderson and May,

1978). While the Feasible k Model predicts that the parasite distribution varies with P and H, the distribution is almost always aggregated (variance / mean > 1, $k < 2$, Fig. 3.1). However, aggregation by itself is not sufficient to stabilize the equilibrium of a host-parasite interaction. Rather aggregation must be an increasing function of mean parasite load (Kretzschmar and Adler, 1993). The $\hat{k}(P, H)$ predicted by the partition model largely satisfies this stability criterion and therefore, given an equilibrium exists for a Feasible k Model and parasites do not affect host fecundity, this equilibrium should be stable.

3.5.2 Stability properties with a parasite reduction in host fecundity ($\xi > 0$)

In the previous section I considered the stability of the host-parasite interaction when there was no parasite-induced reduction in host fecundity ($\xi = 0$). However, in many host-macroparasite systems parasites have larger effects on host fecundity than on host mortality (e.g. Tompkins and Begon, 1999). Therefore, I also consider the stability and dynamics for both the Fixed k and Feasible k Model when $\xi > 0$.

Parasite-induced reduction in host fecundity is a destabilizing mechanism in host-parasite systems (May and Anderson, 1978; Diekmann O. and Kretzschmar, 1991; Dobson and Hudson, 1992). Including parasite-induced reduction in fecundity into the Fixed k Model leads to a region of the parameter space in which the host-macroparasite equilibrium exists, but is unstable (to the left of the stability region in Fig. 3.4). Decreasing k (increasing aggregation) reduces the region in which the equilibrium is unstable (reduces the area of the left hand region in Fig. 3.3), but increases the region in which parasites are unable to regulate the host population (right hand region of Fig. 3.4). Thus

aggregation promotes the stability of the host-macroparasite equilibrium, but decreases the overall parameter space in which an equilibrium can occur. This result is consistent with the similar Fixed k Model explored by Diekmann O. and Kretzschmar (1991).

In comparison to the Fixed k Model where $k \leq 1$ (which is an empirically realistic level of host-macroparasite aggregation, Shaw and Dobson, 1995; Shaw et al., 1998), the Feasible k Model predicts a larger region in which the host-macroparasite has a positive equilibrium that is unstable and, consistent with the results in the previous section, a larger region in which the host population is unregulated (i.e. a finite, positive equilibrium does not exist, Fig. 3.4). This unstable region of the parameter space either showed stable limit cycles or unlimited host growth under the Feasible k Model, after perturbation from the equilibrium (Fig. 3.4). The reason for the increased region of instability in the Feasible k Model compared to the Fixed k Model with $k \leq 1$ is because of the dynamic link between the equilibrium number of hosts/parasites and parasite aggregation. In the Feasible k Model, as parasite pathogenicity (α) decreases, the equilibrium number of hosts also decreases (Fig. 3.3) and k generally increases (aggregation decreases, Fig. 3.3C). Increasing k decreases the stability of the host-macroparasite equilibrium (May and Anderson, 1978) leading to a reduced region in which the host-macroparasite equilibrium is stable.

3.6 Discussion

This study couples constraint-based models of parasite aggregation with dynamic host-macroparasite models to show that when parasites follow empirically- and theoretically-supported predictions of aggregation the ability of a parasite to regulate host populations and stabilize the host-parasite equilibrium is reduced compared to the canonical

assumption of fixed aggregation. These results build on previous studies that show that assuming dynamic parasite aggregation can fundamentally change the behavior of host-macroparasite systems (Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002). However, it has been unclear whether there is a consistent way that real host-macroparasite systems deviate from the assumption of fixed aggregation. Because constraint-based models provide a empirically-supported prediction regarding how parasite aggregation changes with P and H, they may provide a more reasonable starting point when modeling host-macroparasite dynamics than a fixed level of aggregation.

While the Feasible k Model explored here reduced the parameter space in which parasites could regulate host populations, this does not necessarily mean that parasites are less likely to regulate host populations in nature. This is because the reduced parameter space may still be the parameter space in which most host-macroparasite systems reside. In fact, many macroparasites have little effect on the survival probability of their hosts and instead affect other host vital rates such as reproduction (e.g. Dobson and Hudson, 1992; Tompkins and Begon, 1999). Moreover, many macroparasites have high reproductive rates (Anderson and May, 1991), which makes it easier for parasites to regulate host populations under the Feasible k Model. That being said, empirical evidence demonstrating that parasites alone can regulate a host population is uncommon (Scott, 1988; Tompkins et al., 2002, 2011). While this lack of empirical evidence is augmented by the extreme logistical difficulties associated with empirically demonstrating that parasites regulate a host population (Scott and Dobson, 1989; Tompkins et al., 2011), the models presented here suggest that common patterns of parasite aggregation might make the ability of parasites to regulate host populations more rare than previously thought.

In addition to reducing the ability of parasites to regulate a host population, the Feasible k Model, compared to the Fixed k Model, also reduced the stability of the host-

parasite equilibrium when parasites reduced host fecundity. It is well-known that by reducing host fecundity macroparasites can augment cycles in host populations (Dobson and Hudson, 1992; Hudson et al., 1998; Rosà and Pugliese, 2002). The Feasible k Model shows that when parasites reduce the fecundity of their host there is an increased region of the parameter space in which an unstable host-parasite equilibrium exists, compared to the assumptions from the Fixed k Model with k in an empirically realistic range of less than 1 (Shaw et al., 1998). In some of this parameter space, perturbations from the unstable Feasible k Model equilibrium resulted in population cycles when the Fixed k Model predicted a stable host-parasite equilibrium.

The propensity of fixed k models to underestimate the region of host-macroparasite cycles compared to more mechanistically based host-macroparasite models has been previously noted (Rosà and Pugliese, 2002). Rosà and Pugliese showed that fixed k models under-predicted the parameter regions in which oscillatory dynamics were expected if the true mechanism leading to aggregated parasite distributions was clumped infections. However, Rosà and Pugliese (2002) also showed that if heterogeneity in host susceptibility was the aggregating mechanism, then a Fixed k Model over-predicted the region of oscillatory dynamics. The Feasible k Model differs from the extensions explored by Rosà and Pugliese (2002) as the aggregation in the parasite distribution is not directly attributable to a specific process. However, because many different processes result in aggregation patterns that follow the partition model, the systematic under-prediction of oscillatory dynamics by the Fixed k Model may be a general result. On the other hand, both models ignore other well-known destabilizing mechanisms in host-macroparasite systems, such as time delays in the macroparasite life cycle (May and Anderson, 1978), such that a more mechanistic approach to parasite aggregation will be needed to understand how often the Fixed k Model over-predicts or under-predicts the stability of the

host-parasite equilibrium.

This highlights a notable short-coming of the Feasible k Model: it does not directly model the mechanisms leading to parasite aggregation. This is a limitation because it is well-established that many host-parasite mechanisms affect aggregation (Anderson and Gordon, 1982). For example, increasing parasite pathogenicity α can in turn affect the shape of the host-parasite distribution (Crofton, 1971; Barbour and Pugliese, 2000), above and beyond any effects on aggregation via changes in the total number of parasites and the total number of hosts (Johnson and Wilber, 2017). While changes in aggregation independent of changes in P and H do occur in host-parasite systems (Johnson and Wilber, 2017), changes in aggregation are often well-described by changes in P and H (Wilber et al., 2017). Therefore, the assumption that mechanisms affect aggregation primarily through their affects on P and H may not be unreasonable. Moreover, the Feasible k Model assumption is less stringent than the commonly used Fixed k Model (Anderson and May, 1978; Dobson and Hudson, 1992; Townsend et al., 2009). By allowing the level of aggregation to change with P and H, while still maintaining the flexible negative binomial distribution, the Feasible k Model is mimicking a three-dimensional system of equations where the level of parasite aggregation is also included as a dynamic variable (Adler and Krestzschmar, 1992; Kretzschmar and Adler, 1993; Rosà and Pugliese, 2002).

All of these models (i.e. the Fixed k Model, the Feasible k Model, and the three dimensional model) assume that parasites instantaneously redistribute themselves to either maintain the same level of aggregation or change aggregation according to some function. Considering the biological realism of this assumption, it is clearly impossible for a intra-host helminth population to instantaneously redistribute itself following a change in the host or parasite population. While the biological realism and implications of this

assumed instantaneous redistribution of macroparasites is rarely considered in host-macroparasite models, they are much-discussed topics in host-parasitoid systems (e.g. Murdoch and Stewart-Oaten, 1989; Rohani et al., 1994). Depending on the assumptions made about the time scale at which parasitoids redistribute themselves among patches of hosts and whether this redistribution depends on host density, the strength of parasitoid aggregation as a stabilizing mechanism can change (Murdoch and Stewart-Oaten, 1989; Rohani et al., 1994). For example, while density-dependent aggregation is a putative stabilizing mechanisms in discrete time host-parasitoid models (Hassell and May, 1973), Murdoch and Stewart-Oaten (1989) and Rohani et al. (1994) showed that this stabilizing effect was reduced when parasitoids were able to continually redistribute themselves among patches in response to host density within a season. This interaction between the stability properties of a system and the time scale at which putative stabilizing process are assumed to act has been shown in other models as well (e.g. Singh and Nisbet, 2007) and highlights that it is important to identify not just how macroparasite aggregation changes with P and H, but also the time scale at which this change in aggregation occurs after changes in P and H. Incorporating time delays in the response of aggregation to changes in host and parasite abundance is one possible approach to account for this mismatch in time scale. Depending on the nature of the time delay, it could potentially increase or decrease the predicted stability of the host-parasite equilibrium in the Feasible k Model (Nunney, 1985).

These aforementioned challenges stem for the fact that the Feasible k Model, just like the Fixed k Model, is still only approximation of an "exact" model which explicitly tracks the number of hosts with 0, 1, 2, ..., ∞ parasites through time (either deterministically or stochastically, Rosà and Pugliese, 2002; Rosà et al., 2003). Because of this the Feasible k Model will inevitably not be able to capture some of the dynamical properties

that are seen under an exact model. One open question is how much the dynamics of these "exact" models differ from those of the Feasible k Model. Answering this question is challenging as the Feasible k Model implicitly assumes that "many" mechanisms are interacting to lead to emergent patterns of parasite aggregation predicted by the partition model, but it is not entirely clear how to explicitly model these "many" mechanisms. One approach that could prove useful is to construct individual-based models (IBMs) of empirically well-understood host-parasite systems where empirical parasite aggregation follows partition model predictions. These IBMs could be used to explore what combinations of mechanisms are needed to observe emergent patterns of aggregation consistent with the partition model and whether the resulting IBM dynamics are consistent with the Feasible k Model.

In conclusion, variable parasite aggregation can have large effects on the predicted dynamics of host-parasite systems. Constraint-based models provide an empirically and theoretically-supported way to account for the variation in parasite aggregation and show that models that assume fixed aggregation may over-predict both the ability of parasites to regulate a host population and the stability of the host-parasite equilibrium. Given the lack of evidence that macroparasites are the dominant factor regulating many host populations (Tompkins et al., 2011), these results provide a theoretical rationale, to complement the often-discussed logistical rationale (Scott and Dobson, 1989; Tompkins et al., 2011), that may help explain some of these null results.

3.7 Acknowledgments

Thanks to C. Briggs, W. Murdoch, and R. Nisbet for helpful discussions and comments on this manuscript. MW was supported by the National Science Foundation

Graduate Research Fellowship (Grant No.DGE 1144085).

References

Adler, F. R., and M. Krestzschmar. 1992. Aggregation and stability in parasite-host models. Parasitology **104**:199–205.

Albon, S. D., A. Stien, R. J. Irvine, R. Langvatn, E. Ropstad, and O. Halvorsen. 2002. The role of parasites in the dynamics of a reindeer population. Proceedings of the Royal Society B **269**:1625–32.

Anderson, R. M., and D. M. Gordon. 1982. Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. Parasitology **85**:373–398.

Anderson, R. M., and R. M. May. 1978. Regulation and stability of host-parasite interactions: I. Regulatory processes. Journal of Animal Ecology **47**:219–247.

Anderson, R. M., and R. M. May. 1991. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford.

Barbour, A. D., and A. Pugliese. 2000. On the variance-to-mean ratio in models of parasite distributions. Advances in Applied Probability **32**:701–719.

Boag, B., J. Lello, A. Fenton, D. M. Tompkins, and P. J. Hudson. 2001. Patterns of parasite aggregation in the wild European rabbit (*Oryctolagus cuniculus*). International Journal for Parasitology **31**:1421–1428.

Calabrese, J. M., J. L. Brunner, and R. S. Ostfeld. 2011. Partitioning the aggregation of parasites on hosts into intrinsic and extrinsic components via an extended Poisson-gamma mixture model. PloS One **6**:e29215.

Chan, M. S., and V. S. Isham. 1998. A stochastic model of schistosomiasis immuno-epidemiology. Mathematical Biosciences **151**:179–198.

Cornell, S. J., 2010. Modelling stochastic transmission processes in helminth infections. Chapter 5, pages 66–78 *in* E. Michael and R. C. Spear, editors. Modelling Parasite Transmission and Control. Landes Bioscience and Springer Science+Business Media, New York.

Crofton, H. D. 1971. A quantitative approach to parasitism. Parasitology **62**:179–193.

Diekmann O., and M. Kretzschmar. 1991. Patterns in the effects of infectious diseases on population growth. Journal of Mathematical Biology **29**:539–570.

Dobson, A. P., and P. J. Hudson. 1992. Regulation and stability of a free-living host-parasite system: *Trichostrongylus tenuis* in red grouse. II. Population models. Journal of Animal Ecology **61**:487–498.

Fowler, A. C., and T. D. Hollingsworth. 2016. The dynamics of *Ascaris lumbricoides* infections. Bulletin of Mathematical Biology **78**:815 – 833.

Gourbière, S., S. Morand, and D. Waxman. 2015. Fundamental factors determining the nature of parasite aggregation in hosts. PloS One **10**:e0116893.

Gregory, R. D., and M. E. J. Woolhouse. 1993. Quantification of parasite aggregation: a simulation study. Acta Tropica **54**:131–139.

Haegeman, B., and R. S. Etienne. 2010. Entropy maximization and the spatial distribution of species. The American Naturalist **175**:E74–90.

Harte, J. 2011. Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics. Oxford University Press, Oxford, United Kingdom.

Harte, J., and E. A. Newman. 2014. Maximum information entropy: a foundation for ecological theory. Trends in Ecology & Evolution **29**:384–389.

Hassell, M. P., and R. M. May. 1973. Stability in insect host-parasitoid models. Journal of Animal Ecology **42**:693–726.

Hudson, P. J., A. P. Dobson, and D. Newborn. 1998. Prevention of population cycles by parasite removal. Science **282**:2256–2258.

Isham, V. 1995. Stochastic models of host-macroparasite interaction. The Annals of Applied Probability **5**:720–740.

Jaynes, E. 1982. On the rationale of maximum-entropy methods. Proceedings of the IEEE **70**:939–952.

Johnson, P. T. J., and J. T. Hoverman. 2014. Heterogeneous hosts: how variation in host size, behaviour and immunity affects parasite aggregation. The Journal of Animal Ecology **83**:1103–1112.

Johnson, P. T. J., and M. Q. Wilber. 2017. The biological basis of Taylor's Power Law revisited: using host-parasite interactions to reveal the drivers of aggregation. Proceedings of the Royal Society B **284**:20171388.

Jones, E., T. Oliphant, and P. Peterson, 2013. SciPy: Open Source Scientific Tools for Python.

Kretzschmar, M., and F. R. Adler. 1993. Aggregated distributions in models for patchy populations. Theoretical Population Biology **43**:1–30.

Locey, K. J., and D. J. McGlinn. 2013. Efficient algorithms for sampling feasible sets of macroecological patterns. PeerJ pages 1–23.

Locey, K. J., and E. P. White. 2013. How species richness and total abundance constrain the distribution of abundance. Ecology Letters **16**:1177–85.

May, R. M., and R. M. Anderson. 1978. Regulation and stability of host-parasite population interactions: II. Destabilizing processes. Journal of Animal Ecology **47**:249–267.

Murdoch, W. W., J. Chesson, and P. L. Chesson. 1985. Biological control in theory and practice. The American Naturalist **125**:344–366.

Murdoch, W. W., and A. Stewart-Oaten. 1989. Aggregation by parasitoids and predators: effects on equilibrium and stability. The American Naturalist **134**:288–310.

Nisbet, R. M., and W. S. C. Gurney. 1982. Modelling Fluctuating Populations. John Wiley and Sons, New York.

Nunney, L. 1985. Short time delays in population models: A role in enhancing stability. Ecology **66**:1849–1858.

Pacala, S. W., and A. P. Dobson. 1988. The relation between the number of parasites/host and host age: population dynamic causes and maximum likelihood estimation. Parasitology **96**:197–210.

Peterson, M. J. 2004. Parasites and infectious diseases of prairie grouse: should managers be concerned? Wildlife Society Bulletin **32**:35–55.

Pugliese, A., and R. Rosa. 1995. A 2-dimensional model for macroparasitic infections in a host with logistic growth. Journal of Biological Systems **3**:833–849.

Pugliese, A., R. Rosà, and M. L. Damaggio. 1998. Analysis of model for macroparasitic infection with variable aggregation and clumped infections. Journal of Mathematical Biology **36**:419–47.

Redpath, S. M., F. Mougeot, F. M. Leckie, D. A. Elston, and P. J. Hudson. 2006. Testing the role of parasites in driving the cyclic population dynamics of a gamebird. Ecology Letters **9**:410–418.

Rohani, P., H. C. J. Godfray, and M. P. Hassell. 1994. Aggregation and the dynamics of host-parasitoid systems: a discrete-generation model with within-generation redistribution. The American Naturalist **144**:491–509.

Rosà, R., and A. Pugliese. 2002. Aggregation, stability, and oscillations in different models for host-macroparasite interactions. Theoretical Population Biology **61**:319–34.

Rosà, R., A. Pugliese, A. Villani, and A. Rizzoli. 2003. Individual-based vs. deterministic models for macroparasites: host cycles and extinction. Theoretical Population Biology **63**:295–307.

Scott, M. E. 1987. Temporal changes in aggregation: a laboratory study. Parasitology **94**:583–595.

Scott, M. E. 1988. The impact of infection and disease on animal populations: implications for conservation biology. Conservation Biology **2**:40–56.

Scott, M. E., and A. P. Dobson. 1989. The role of parasites in regulating host abundance. Parasitology Today **5**:176–183.

Shaw, D. J., and A. P. Dobson. 1995. Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. Parasitology **111**:111–133.

Shaw, D. J., B. T. Grenfell, and A. P. Dobson. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology **117**:597–610.

Singh, A., W. W. Murdoch, and R. M. Nisbet. 2009. Skewed attacks, stability, and host suppression. Ecology **90**:1679–1686.

Singh, A., and R. M. Nisbet. 2007. Semi-discrete host-parasitoid models. Journal of Theoretical Biology **247**:733–742.

Tompkins, D. M., and M. Begon. 1999. Parasites can regulate wildlife populations. Parasitology Today **15**:311–313.

Tompkins, D. M., A. P. Dobson, P. Arneberg, M. Begon, I. M. Cattadori, J. V. Greenman, J. A. P. Heesterbeek, P. J. Hudson, D. Newborn, A. Pugliese, A. P. Rizzoli, R. Rosa, F. Rosso, and K. Wilson, 2002. Parasites and host population dynamics. Chapter 3, pages 45–62 *in* P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heessterbeck, and A. P. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Tompkins, D. M., A. M. Dunn, M. J. Smith, and S. Telfer. 2011. Wildlife diseases: from individuals to ecosystems. Journal of Animal Ecology **80**:19–38.

Townsend, S. E., S. Newey, S. J. Thirgood, L. Matthews, and D. T. Haydon. 2009.

Can parasites drive population cycles in mountain hares? Proceedings of the Royal Society B **276**:1611–7.

Wilber, M., P. T. J. Johnson, and C. J. Briggs. 2017. When can we infer mechanism from parasite aggregation? A constraint-based approach to disease ecology. Ecology **98**:688–702.

Wilson, K., O. N. Bjoernstad, A. P. Dobson, S. Merler, G. Poglayen, A. F. Read, and A. Skorping, 2002. Heterogeneities in macroparasite infections: patterns and processes. Chapter 2, pages 6–44 *in* P. J. Hudson, A. Rizzoli, B. Grenfell, H. Heesterbeek, and A. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Xiao, X., K. J. Locey, and E. P. White. 2015. A process-independent explanation for the general form of Taylor's Law. The American Naturalist **186**:E51–60.

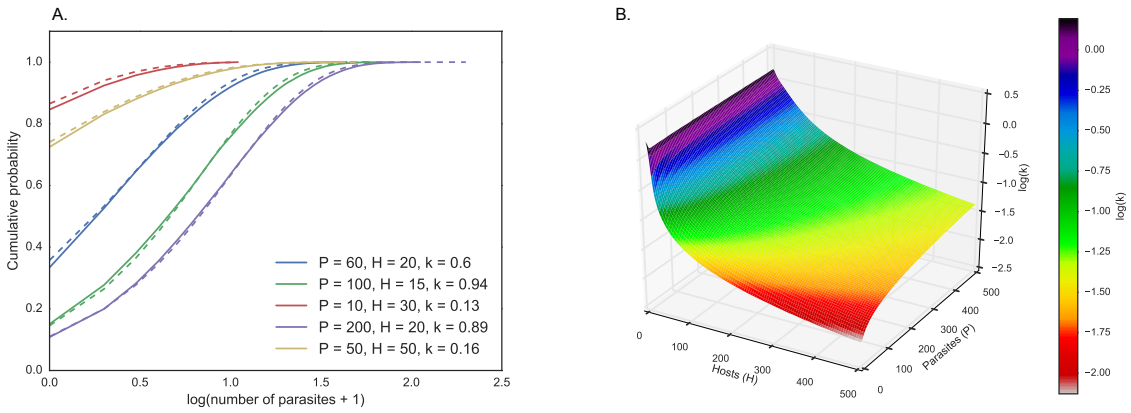


Figure 3.1: **A.** Predicted parasite distributions from the partition model with different numbers of parasites P and hosts H (solid lines) The corresponding best-fit negative binomial distributions are also plotted (dashed lines). The negative binomial distribution with varying levels of k can generally capture the aggregation predictions of the partition model. **B.** The surface $\hat{k}(P, H)$ as predicted by the partition model after interpolation and smoothing. For visual clarity, B. shows $\hat{k}(P, H)$ interpolated over 1000 x 1000 points with H and P between 3 and 500 rather than the 10,000 x 10,000 points used in the analyses.

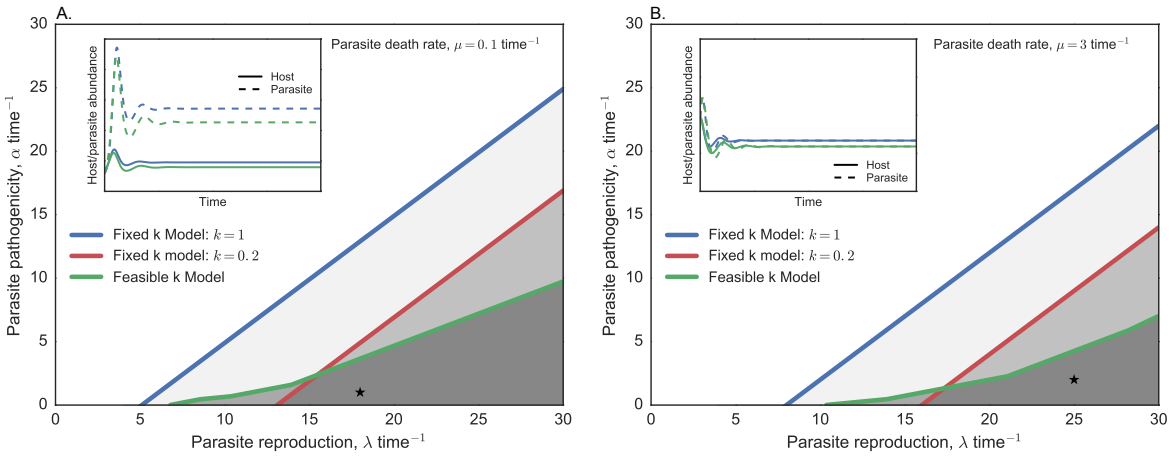


Figure 3.2: **A. - B.** The boundaries at which a parasite can regulate a host population under the Fixed k Model and the Feasible k Model. Below the line of a particular model is where a parasite can regulate a host population. For the parameter space shown, when a parasite was able to regulate the host population the system had a locally stable equilibrium (the host-parasite equilibriums in all gray shaded regions are stable). The inset plots in A. and B. give examples of the host-parasite dynamics when the models have values of λ and α given by the black star in the respective plots. For clarity, the inset plots do not show the dynamical predictions for the Fixed k Model with $k = 0.2$. All other parameters are $\mu = 0.1$ and 3 time^{-1}, $H_0 = 10$, $b = 3$ time^{-1}, $d = 1$ time^{-1}, and $\xi = 0$.

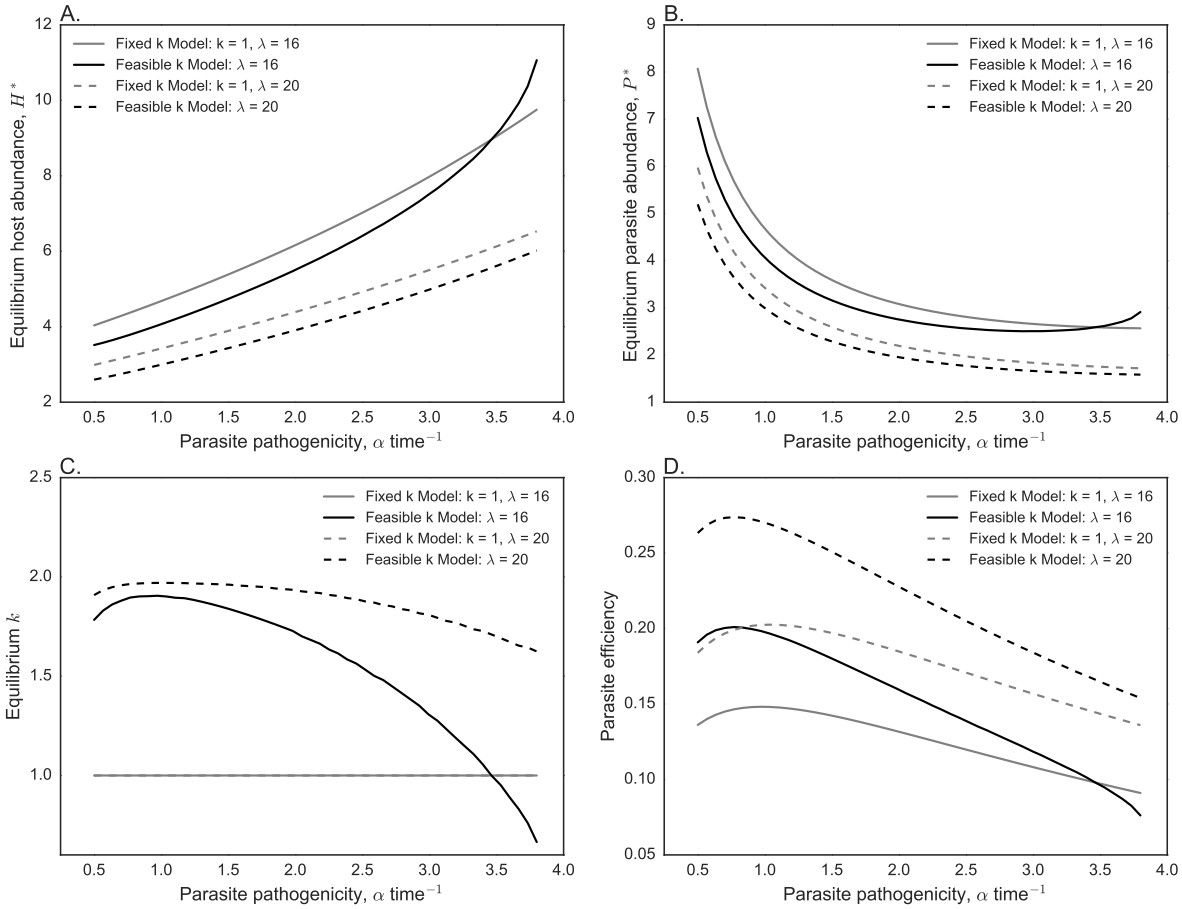


Figure 3.3: **A.** The host equilibrium under the Fixed k Model and the Feasible k Model with different values of parasite reproduction λ and varying levels of parasite pathogenicity α. **B.** The parasite equilibrium under the Fixed k and Feasible k Models. **C.** The equilibrium level of k under the Fixed k and Feasible k Models. The k for the Fixed k Model, by definition, does not change. **D.** Parasite efficiency, defined as the negative log percentage of unparasitized hosts divided by equilibrium parasite abundance, under the Fixed k and Feasible k Models. The other parameters are $\mu = 0.1$ time^{-1}, $H_0 = 10$, $b = 3$ time^{-1}, $d = 2$ time^{-1}, and $\xi = 0$.

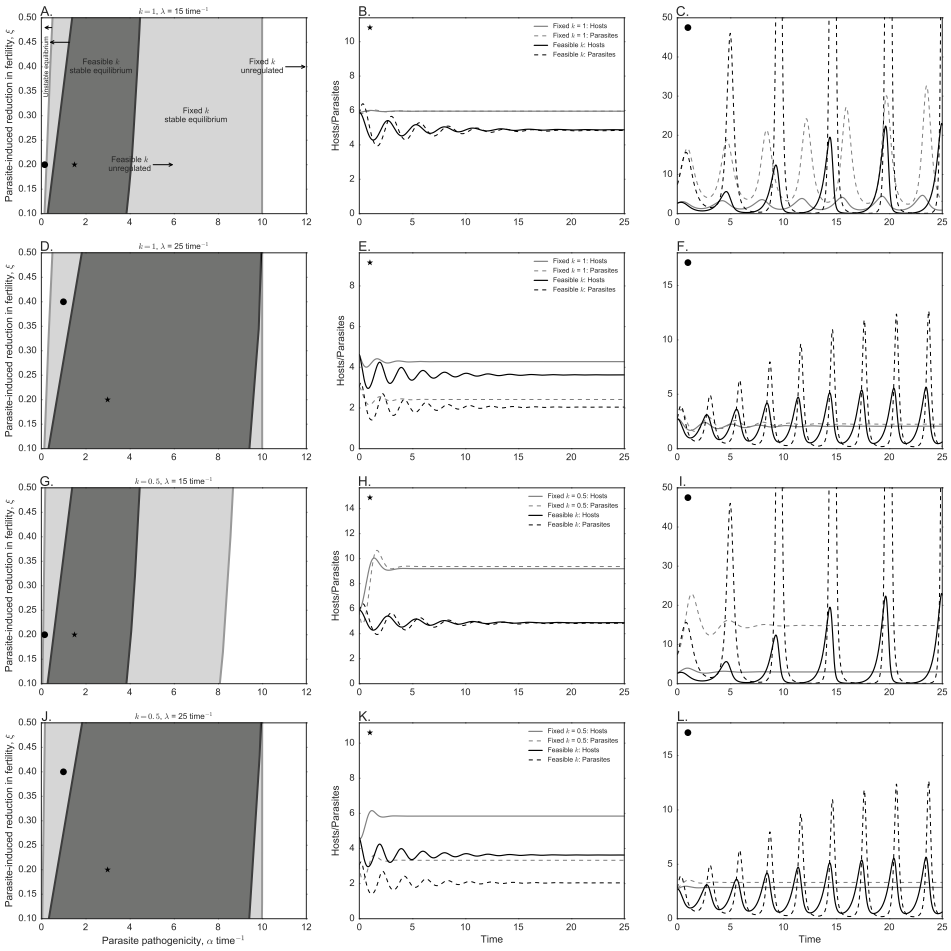


Figure 3.4: Stability properties and dynamics of the Fixed k and Feasible k Models when parasites cause a multiplicative, load-dependent reduction in host fecundity. **A., D., G., J.** The parameter space in which the host-parasite equilibrium is stable for varying parasite pathogenicity α, parasite reduction in fecundity ξ, parasite reproduction $\lambda = 10$ time^{-1}, and k for the Feasible k and Fixed k Model. The dark gray regions show where the Feasible k Model has a stable equilibrium and the light gray regions show where the Fixed k Model has a stable equilibrium. To the right of the shaded regions the parasite fails to regulate the host population. To the left of the shaded regions the parasite can regulate the host, but the resulting equilibrium is unstable. The dynamics from these different regions are shown in **B., C., E., F., H., I., K., L.** which correspond to the black circles and stars shown in **A., D., G., J.**. The y-axis is truncated on **C.** and **I.** for clarity. All other parameters are $\mu = 0.1$ time^{-1}, $H_0 = 10$, $b = 3$ time^{-1}, and $d = 1$ time^{-1}.

Chapter 4

Integral Projection Models for host-parasite systems with an application to amphibian chytrid fungus

4.1 Abstract

1. Host parasite models are typically constructed under either a microparasite or macroparasite paradigm. However, this has long been recognized as a false dichotomy because many infectious disease agents, including most fungal pathogens, have attributes of both microparasites and macroparasites.

2. We illustrate how Integral Projection Models (IPM)s provide a novel, elegant modeling framework to represent both types of pathogens. We build a simple host-parasite IPM that tracks both the number of susceptible and infected hosts and the distribution of parasite burdens in infected hosts.

3. The vital rate functions necessary to build IPMs for disease dynamics share many commonalities with classic micro and macroparasite models and we discuss how these functions can be parameterized to build a host-parasite IPM. We illustrate the utility of this IPM approach by modeling the temperature-dependent epizootic dynamics of amphibian chytrid fungus in Mountain yellow-legged frogs (*Rana muscosa*).

4. The host-parasite IPM can be applied to other diseases such as facial tumor disease in Tasmanian devils and white-nose syndrome in bats. Moreover, the host-parasite IPM can be easily extended to capture more complex disease dynamics and provides an exciting new frontier in modeling wildlife disease.

4.2 Introduction

Following the influential papers by Anderson and May (Anderson and May, 1979; May and Anderson, 1979), host parasite models have usually been constructed within one of two model structures. In their simplest form, microparasite models classify individuals

as susceptible, infected or recovered (SIR), with the implicit assumption that all infected hosts can be considered similar because once a host is infected microparasites can rapidly multiply within the host. Under this simple structure, prevalence, the proportion of infected individuals, is therefore adequate to characterize the level of infection within a host population. In contrast, macroparasite models generally assume that parasites cannot complete their entire life cycle within an individual host. Therefore, infection levels within a host are strongly influenced by the number of infective stages the host has encountered, and parasite burden influences host survival, reproduction and the transmission of infective stages. As a result, in macroparasite models, the proportion of individuals infected is not adequate to characterize the level of infection within a host population, and therefore it is necessary to model the frequency distributions of parasites among individuals.

In some pathogens traditionally categorized as microparasites, pathogen within-host reproduction occurs at a slow enough rate that it can be tracked from one time point to the next (e.g. Briggs et al., 2010; Langwig et al., 2015*a*). In these instances, it is useful to take a macroparasite approach and model the distribution of loads across hosts as this measure is both more consistent with the type of data collected on these diseases and allows for the prediction of additional epidemiological patterns such as the dynamics of parasite aggregation (Scott, 1987). Fungal pathogens are increasingly recognized as important threats to biodiversity, agricultural production and human health (Fisher et al., 2012) and may exhibit this relatively slow, measurable on-host reproduction. A modeling framework that accounts for both their microparasite and macroparasite characteristics is critical for understanding their dynamics.

To this end, Briggs et al. (2010) developed an individual based model for the fungal pathogen *Batrachochytrium dendrobatidis* in frog populations and were able to predict

the biological criteria necessary for population persistence as well as the efficacy of different treatment strategies during epizootics (Drawert et al., 2015). However, this model required a separate equation for the fungal load on each individual and was difficult to parameterize from field or experimental data. In general, there is a need for an intermediate modeling framework for "slow" microparasites that accounts for the information in the distribution of parasites across hosts, while allowing for straightforward parameterization from laboratory or field data.

In this paper, we illustrate the potential for integral projection models (IPM)s to address this need. Several recent papers have provided excellent overviews of the construction and use of IPMs (Rees and Ellner, 2009; Coulson, 2012; Metcalf et al., 2013; Merow et al., 2014*a*; Rees et al., 2014; Metcalf et al., 2015). In very general terms, IPMs assume that demographic parameters of individuals are affected by one or more continuous variables that describe some property of those individuals. The models then iterate population dynamics in discrete time with state variables of the form $N(x,t)$, representing the frequency of individuals with continuous property x at the time t.

The models can be readily parameterized from data using linear or non-linear regression based approaches. For population models, the continuous variable x is often the size, such as body mass (e.g. Coulson, 2012), or age of an organism, but in principle any continuous variable or variables could be used. Here we illustrate their use, using a measure of parasite load. It has been pointed out that this approach may well be suitable for modeling host parasite interactions (Cooch et al., 2012; Metcalf et al., 2015), but we know of only one application of these models for wildlife pathogens: a model of the fungal infection aspergillosis in sea fans (Bruno et al., 2011). The model in that study categorized sea fans into either infected or uninfected categories, and the continuous variable modeled by the integral projection approach was the size of the sea

fan, and not the parasite load itself. Recently, Metcalf et al. (2015) have proposed a general framework for using IPM models for disease in which they highlight some of the benefits and challenges of fitting disease data to these models. Here we build on the ideas proposed by Metcalf et al. by providing a detailed case study and other examples of how these methods could be used to address key questions in disease ecology and evolution. Where possible, we try to use similar notation as Metcalf et al..

4.3 Materials and Methods

The basic model we examine is a modification of a susceptible-infected-susceptible (SIS) model. In our model, individuals that clear the infection immediately re-enter the susceptible class, with no immunity. Including a recovered class simply requires adding an additional discrete stage to the IPM (Metcalf et al., 2015). The model has the following state variables:

$S(t)$: Number of susceptible/uninfected hosts at time t

$I(x,t)$: Frequency of infected hosts with load x at time t (where $x \neq 0$). i.e. The number of hosts at time t with a load between lower bound (L) and upper bound (U) is $\int_L^U I(x,t)dx$

In a traditional macroparasite model x is an integer, so the integral could be replaced with a summation. However, an advantage of the IPM approach is that regression techniques can be used to parameterize the model, rather than needing to estimate a large number of individual matrix parameters. For most fungal infections and many other parasites, data on infection intensity are generally measured using quantitative

PCR (Boyle et al., 2004), and continuous measures of infection load are more appropriate than discrete counts.

Following Rees et al. (2014), the system can be represented by the life history flow chart in Figure 4.1. This can be written as the following equation for susceptible/uninfected hosts at time $t+1$

$$S(t+1) = S(t)s_0(1-\phi(I(x,t))+$$

$$\int_L^U I(x,t)s(x)l(x)dx + \left(f_0 S(t) + \int_L^U f(x)I(x,t)dx\right) \tag{4.1}$$

The first term in this equation gives the number of hosts who remain uninfected in a time step. The second term gives the number of infected hosts who lose an infection and enter the uninfected class in a time step. The third term gives the number of uninfected hosts who are born from uninfected hosts in a time step. The fourth terms give the number of uninfected hosts that are born from infected hosts in a time step.

The equation for infected hosts with load x' at time $t+1$ is given by

$$I(x',t+1) = \int_L^U I(x,t)s(x)(1-l(x))G(x',x)dx + S(t)s_0\phi(I(x,t))G_0(x') \tag{4.2}$$

The first term in this equation gives the number of infected individuals with load x that transition to load x' in a time step. The second term gives the number of uninfected individuals that transition to an infected individual with load x' in a time step. Below we more thoroughly discuss the terms in equations 1 and 2, how they relate to classic macroparasite and microparasite models, and how they can be parameterized. When parameterizing the functions below, we assume that each process obeys the Markov

property such that only the load at time t predicts the event at time $t+1$ (e.g. growth of the parasite, host survival, loss of infection, etc.) (Easterling et al., 2000).

4.3.1 The growth function: $G(x', x)$

For continuous measures of parasite load, the growth function $G(x', x)$ specifies the probability density of transitioning to load x' at time $t+1$, dependent on having a load of x at time t. In comparison to standard macroparasite and microparasite models, this function allows for pathogen growth on a host to be driven by both within host pathogen birth/rapid self-reinfection (e.g. microparasites and some macroparasites) and from acquiring additional parasites from the environment or other infected hosts. The dependence of $G(x', x)$ on the free-living stages of the parasites can be made explicit by writing $G(x', x)$ as dependent on the number of free-living parasites at time t.

This function can be estimated with data on the parasite load of individual hosts at time t and time $t+1$. Using standard regression techniques, load at time t, the number of free-living parasites at time t, and/or the density and abundance of other infected hosts can be regressed against load at time $t+1$ and the resulting model can be used to parametrize the growth function $G(x', x)$ (Easterling et al., 2000). For continuous parasite loads, the load at time $t+1$ could be described by a lognormal or gamma distribution, while discrete disease loads could be fit by a negative binomial distribution (Anderson and May, 1978; Shaw et al., 1998). The growth of a parasite on a host will often depend on other abiotic variables that can be accounted for as additional fixed or random effects in the regression model (Rees and Ellner, 2009).

4.3.2 The survival function: s_0 and $s(x)$

s_0 specifies the survival probability of uninfected hosts. $1 - s_0$ gives the probability of a host dying without any infection, which parallels the death rate of uninfected hosts in classic micro and macroparasite models. s_0 can be estimated by the proportion of uninfected hosts that survive from t to $t + 1$.

The survival function $s(x)$ specifies the probability of a host with a parasite load x surviving from time t to time $t + 1$. In classic macroparasite models, it is assumed that parasite-induced host mortality increases linearly with load at rate α, where α specifies the pathogenicity of the parasite (Anderson and May, 1978). In the IPM framework, a commonly used function to measure survival probability is the logistic function given by

$$s(x) = \frac{\exp(b_0 - b_1 x)}{1 + \exp(b_0 - b_1 x)} \tag{4.3}$$

where b_1 is similar to the pathogenicity parameter α (Anderson and May, 1978; Wilber et al., 2016). When b_1 is held constant, b_0 dictates the parasite load at which substantial parasite-induced host mortality begins to occur (Wilber et al., 2016). The logistic function could be replaced by other functions, as dictated by the data (Dahlgren et al., 2011).

The survival function $s(x)$ can be estimated with logistic regression using host survival and load data from laboratory or mark-recapture studies conducted at the appropriate time scale. If other biotic or abiotic factors are also thought to contribute to the survival probability of a host from t to $t+1$, they can be included as additional predictor variables in the survival function.

4.3.3 The loss of infection function: $l(x)$

The loss of infection function $l(x)$ specifies the probability of a host having a parasite load of x at time t and losing the infection by time $t+1$. In comparison with classic microparasite susceptible-infected models, this function is analogous to the rate at which infected individuals recover from infection and reenter the susceptible/uninfected class. We similarly assume that individuals that lose an infection immediately reenter the susceptible/uninfected class, though a resistant class could easily be included in this modeling framework (Metcalf et al., 2015).

A logistic function (equation 3) could also be used for the loss of infection function and could be parameterized using parasite load data at time t and $t+1$ and fitting a logistic regression where the response variable is whether or not a host lost an infection by time $t+1$ and the predictor variable is the infection intensity x at time t. As with the survival function, if other biotic or abiotic factors also contributed to $l(x)$ they could be included as additional predictor variables in the logistic regression.

4.3.4 The transmission function: $\phi(I(x,t))$

The transmission function $\phi(I(x,t))$ specifies the probability of transitioning from the uninfected class to the infected class. The transmission function is critically important for the dynamics of a disease and can take a variety of different functional forms. Some common examples include the density-dependent or mass action transmission function βIS and the frequency-dependent transmission function $\beta SI/N$ (McCallum et al., 2001).

Over a unit time interval, a density-dependent, mass action transmission function results in the following probability of an individual host not being infected:
$\exp\left(-\beta \int_L^U I(x,t)dx\right)$, where $\int_L^U I(x,t)dx$ gives the total number of infected individuals.

Allowing β to be a function of parasite load, $\phi(I(x,t))$ can be written as

$$\phi(I(x,t)) = 1 - \exp\left(-\int_L^U \beta(x)I(x,t)dx\right) \tag{4.4}$$

where $\int_L^U \beta(x)I(x,t)dx$ is the force of infection and $\beta(x)$ specifies the effect of an individual with an infection load of x on the infection probability of an uninfected individual. This formulation assumes that new infections occur following a Poisson process with rate $\int_L^U \beta(x)I(x,t)dx$.

While this functional form may be appropriate for many microparasites in which direct transmission among hosts is the primary mode of acquiring infection, the transmission of some pathogens depends on the number of free-living parasites in a system as well as the number of infected hosts (Briggs et al., 2010). If we assume that number of free-living parasites is proportional to the total number of parasites in all infected hosts in the system at time t, then we can modify $\beta(x)I(x,t)$ to $\beta(x)xI(x,t)$ to capture this biology.

Finally, some pathogens have an environmental reservoir such that the probability of infection is non-zero even when no infected hosts are present. This could be captured by rewriting equation 4 as

$$\phi(I(x,t)) = 1 - \exp\left(-(a + \int_L^U \beta(x)I(x,t)dx)\right) \tag{4.5}$$

where $1-\exp(-a)$ defines the probability of infection when no infected hosts are present (e.g. from an environmental reservoir). This environmental reservoir could be more explicitly accounted for by including an additional state variable in the IPM that tracks how the number of parasites in the environment grows and decays in a time step (Rohani et al., 2009).

Methods for estimating the transmission function and/or its corresponding parameters are well-described in the host-pathogen literature (e.g. McCallum, 2000; Smith et al., 2009), though choosing between transmission functions is typically a data-intensive procedure (Rachowicz and Briggs, 2007; Smith et al., 2009).

4.3.5 The initial infection burden function: $G_0(x')$

The function $G_0(x')$ specifies the probability density of the infection intensity of a host when it first becomes infected and can be a function of the total number of infected hosts in the population, the total number of infectious agents in the population and/or various other host or abiotic covariates. This function can be estimated by fitting a regression model where the response variable is the pathogen load of infected hosts at time $t + 1$ that were uninfected at time t. For continuous disease loads, a variety of different distributions such as gamma, lognormal, and normal could be explored.

In comparison with standard stochastic macroparasite models, this function is analogous to clumped infection distributions in which a host can acquire a random number of free-living parasites in a small time step (Isham, 1995; Pugliese et al., 1998). However, depending on the time step used to parameterize the IPM, $G_0(x')$ will also be influenced by the growth of the parasite on the host as a "clump" of parasites can infect a host and then grow in the time interval t to $t + 1$.

Moreover, the above host-parasite IPM assumes that after acquiring an initial infection burden the growth of the parasite on a host is then driven by $G(x', x)$ and is independent of the density of infected hosts. If one has reason to believe that transmission and the function $G_0(x')$ are important drivers of disease dynamics on hosts after

the initial infection, the growth function may be redefined as

$$G'(x', x, I(x,t)) = (1 - \phi(I(x,t))G(x', x) + \phi(I(x,t)\left[G_0(x' - x) + \text{higher order terms}\right] \tag{4.6}$$

where an increase in load from x to x' in a time step could be because of 1) no transmission occurring and parasite load increasing due to within host growth (first term) or 2) transmission occurring and a host acquiring a "clump" of infections of size y such that $y = x' - x$ (second term) or 3) some combination of both within host growth and transmission occurring such that parasite load increases from x to x' in a time step. This is given by higher order terms and will depend on the length of the time step t to $t + 1$ relative to the dynamics of the pathogen.

4.3.6 The fecundity function: f_0 and $f(x)$

The fecundity function $f(x)$ specifies the mean number of offspring produced by individuals with a parasite load of x (or by susceptible/uninfected individuals f_0) and the host-parasite IPM assumes that all offspring enter the uninfected class. It is easy to relax this assumption and include vertical transmission into the host-parasite IPM by allowing newly born hosts to enter the infected class with a parasite load specified by some probability density function. Standard macroparasite models assume that host reproduction decreases linearly with increasing parasite load (May and Anderson, 1978). However, as pointed out by May and Anderson, this is an over simplification as the response of host reproductive effort to parasitism is often non-linear (e.g. Weatherly, 1971) and reproduction itself can never take on a negative value (Roberts et al., 1995). Alternative formulations of parasite-induced reduction in host fertility that account for

this non-linear relationship have been discussed (Roberts et al., 1995).

In the IPM framework, the fecundity function can be fit using Poisson or negative binomial regression where the predictor variable is parasite load and the response variable is the number of offspring produced by a host with that parasite load (Easterling et al., 2000). If the response variable is a non-integer value, a continuous distribution such as gamma or lognormal could be used. Depending on the link function, these regressions can produce non-linear fecundity functions that are always positive. Similar to the other vital functions discussed above, other factors that affect mean reproductive output can be included in the regression.

For many host-parasite systems, host reproduction occurs on a much longer time scale than the dynamics of the parasite and it may not be biologically realistic to include host reproduction at each time step in the IPM model as is done in equation 4.1. For example, if host reproduction occurs at a particular time during the year it may be useful to break the year into separate IPMs (e.g. an IPM for summer, fall, winter and spring; Caswell, 2001) such that load-dependent host reproduction only occurs in a particular season or as a discrete pulse at the beginning of a particular season (e.g. host reproduction is only non-zero in the spring and fall). One may also want to include host age as an additional discrete or continuous host attribute (Childs et al., 2003) to account for reproductive differences among hosts of different ages. On the other hand, if one is particularly interested in the fate of a host population over a single seasonal epizootic where host reproduction does not occur, the fecundity function may be excluded from the host-parasite IPM as it will not affect host population persistence during the epizootic. In this case, appropriately modeling vital rates such as the survival function $s(x)$ and the growth function $G(x', x)$ will be critically important for understanding host-parasite dynamics. In general, how to include host reproduction into the host-parasite IPM will

depend on the questions that are being addressed.

4.3.7 Application of model to amphibian chytrid fungus: Laboratory experiment

We use the above IPM framework to examine the population dynamics of amphibian hosts infected with the fungal pathogen *Batrachochytrium dendrobatidis* (*Bd*). *Bd* is a devastating amphibian pathogen that has led to declines in many amphibian populations around the globe (Skerratt et al., 2007; Kilpatrick et al., 2010). *Bd* is a cutaneous fungus that disrupts the osmoregulatory ability of amphibian skin, eventually leading to chytridiomycosis and amphibian mortality (Voyles et al., 2007, 2009). In contrast to traditional macroparasities, *Bd* rapidly reinfects an infected host in a process analogous to within host birth (Rollins-Smith, 2009). The generation time of *Bd* is between four to ten days depending on temperature (Woodhams et al., 2008), such that the on-host *Bd* growth dynamics can be captured via repeated swabbing of an animal every few days, with the fungal load on the frog estimated as the number of copies of *Bd* DNA detected on the skin swabs via quantitative PCR (Boyle et al., 2004). Quantitative PCR provides a continuous measure of infection intensity between 0 (uninfected) and an arbitrarily large *Bd* infection. These characteristics of *Bd* make it an ideal candidate for applying the host-parasite IPM described above.

We use the IPM framework to gain insight into how temperature affects the epizootic dynamics of *Bd* in populations of the Mountain yellow-legged frog complex (*Rana muscosa* and *Rana sierrae*, henceforth *R. muscosa*). *Rana muscosa* are native to the California Sierra Nevada mountains and have suffered severe *Bd*-induced population declines (Vredenburg et al., 2010; Briggs et al., 2010). The severity of *Bd* infection is

highly temperature dependent (Berger et al., 2004; Andre et al., 2008), with optimal *Bd* growth occurring between 17 - 25 °C in laboratory conditions (Piotrowski et al., 2004), but depending on host-*Bd* interactions (Piotrowski et al., 2004; Raffel et al., 2012). While these are the temperatures at which amphibians often suffer more severe chytridiomycosis and mortality, this pathology is species-dependent (Kilpatrick et al., 2010).

We use data from a laboratory experiment in which 20 adult *R. muscosa* were housed separately at 4 different temperatures (4 °C, 12 °C, 20 °C, 26 °C; 5 frogs per temperature), exposed to approximately 10^6 zoospores of *Bd* and then monitored for 119 days. Every three days starting eight days after exposing the frogs to *Bd*, the frogs were swabbed and *Bd* zoospore load was estimated using quantitative PCR. Mortality that occurred between swabbing events was recorded at the next swabbing event. Frogs housed at 26 °C were visibly distressed and suffered much higher *Bd*-independent mortality than those housed at lower temperatures. For this reason, and because we wished to examine how temperature affected *R. muscosa*-*Bd* dynamics at the much cooler temperatures typically observed in the field (4-20 °C, Knapp et al., 2011), we excluded individuals at 26 °C.

4.3.8 Model description

To fit the IPM to *Bd* load data from laboratory experiments, we made two simplifying assumptions. First, we excluded reproduction/recruitment because we lack data on the effect of infection on reproduction. As a result, we used this model to address questions regarding epizootic dynamics of *Bd* and *R. muscosa* over the course of a single summer season, rather than to examine long-term population persistence with disease.

Second, we assumed the probability of infection $\phi(T)$ was temperature (T)-dependent, but independent of the density of infected hosts (i.e. $I(x,t)$ does not affect the probability of infection). In our experiments individual animals were housed in separate containers and initial infection was solely due to an amphibian acquiring *Bd* zoospores from the environment. We subsequently explore different transmission functions that do include $I(x,t)$ to understand their implications on *Bd* epizootic dynamics. With these assumptions, the modified IPM is given by

$$S(t+1) = S(t)s_0(T)(1-\phi(T)) + \int_L^U I(x,t)s(x,T)l(x,T)dx \tag{4.7}$$

$$I(x',t+1) = \int_L^U I(x,t)s(x,T)(1-l(x,T))G(x',x,T)dx + S(t)s_0(T)\phi(T)G_0(x',T) \tag{4.8}$$

where the various vital functions are now dependent on temperature T. Note that x refers to $\ln(x)$ (log zoospore load) when $x \neq 0$ and 0 (uninfected) when a frog was uninfected. In this case, 0/uninfected represents a discrete state of the frog and is not equivalent to $\ln(x) = 0$. In this model, a single time step represents three days, which was the time between swabbing events in the laboratory experiment.

4.3.9 Vital rate functions

We modeled the survival function $s(x)$ of a frog with a log zoospore load of x as a logistic regression with the link function given by

$$\text{logit}(s(x)) = b_{0,0} + b_{1,0}x \tag{4.9}$$

where $b_{0,0}$ is the intercept of the link function on the logit scale, and $b_{1,0}$ is the effect of log zoospore load on the logit-transformed probability of survival. We could not estimate

the effect of temperature on this vital rate function because no individuals died at 4 °C or 12 °C (Fig. 4.2A). This result was surprising because individuals at 12 °C had loads as high or higher than individuals at 20 °C and did not experience mortality. Based on previous results in the field which show that *R. muscosa* suffer a roughly consistent *Bd*-induced mortality across a variable lake temperatures in the field (i.e. the $\approx$ 10,000 zoospore threshold, Vredenburg et al., 2010), additional results in the laboratory that show that frogs experience significant *Bd*-induced mortality at temperatures below 20 °C (17 °C, Andre et al., 2008), and extensive field observations that decreased temperature does not have a large protective effect on *R. muscosa* in the field (Knapp et al., 2011), we think there is very little evidence that the survival curve of *R. muscosa* and *Bd*-load interacts with temperature. Therefore, we assumed that *Bd*-induced mortality is dependent only on load and not on temperature directly. We therefore parameterized the survival function using only individuals at 20 °C (Fig. 4.2A, see Section 4.A for a comparison with a survival function fitted with all of the temperature data), but assumed a temperature-independent survival function. However, temperature influenced fungal growth, as detailed next.

We modeled the growth function $G(x', x)$ as a normal distribution $X \sim N(\mu(x, T), \sigma^2(x))$ where T is temperature. Mean fungal loads were modeled as

$$\mu(x, T) = b_{0,1} + b_{1,1}x + b_{2,1}T \tag{4.10}$$

where $b_{0,1}$ is the intercept and $b_{1,1}$ and $b_{2,1}$ give the effect of a unit change in log zoospore load and temperature on the log zoospore load at time $t+1$, respectively. We also allowed the variance of $G(x', x)$ to be an exponential function of log zoospore load at time t

$$\sigma^2(x) = \nu_{0,1} \exp(2c_{0,1}x) \tag{4.11}$$

where $\nu_{0,1}$ is a constant and $c_{0,1}$ dictates the effect of log zoospore load on the variance.

We modeled the loss of infection function $l(x)$ as a logistic regression with the link function

$$\text{logit}(l(x,T)) = b_{0,2} + b_{1,2}x + b_{2,2}T \tag{4.12}$$

where $b_{0,2}$ is the intercept and $b_{1,2}$ and $b_{2,2}$ are the coefficients giving the effect of a unit change in log zoopsore load and temperature on the logit-transformed probability of losing an infection in a single time step, respectively.

We modeled the initial infection burden function $G_0(x')$ as a normal distribution $X \sim N(\mu(T), \sigma^2(T))$. We defined the mean of the distribution as $\mu(T) = b_{0,3} + b_{1,3}T$ where $b_{0,3}$ and $b_{1,3}$ are defined similarly to the growth function. We modeled the variance as $\sigma^2(T) = \nu_{0,3}\exp(2c_{0,3}T)$ where $\nu_{0,3}$ and $c_{0,3}$ are defined similarly as in the growth function.

Finally, we modeled the probability of an individual becoming infected $\phi(T)$ in a time step as a function of temperature T using a logistic model $\text{logit}[\phi(T)] = b_{0,4} + b_{1,4}T$ where $b_{0,4}$ and $b_{1,4}$ are defined similar to the recovery function. We performed model selection and validation various for each vital rate functions described above and these results are given in Section 4.A. We fit the vital rate functions in R version 3.1.2 using the functions gls, lme, and glm and all code used for this analysis can be found at `https://github.com/mqwilber/ipm_for_parasites`.

4.3.10 Analyzing the IPM

After fitting the parameters of the vital rate functions, we analyzed the resulting IPM (equation 7) by discretizing the continuous variable *Bd* load and using the midpoint rule

to evaluate the IPM at each time step (Rees et al., 2014). For the infected portion of the host-parasite IPM, we used 100 discretized bins (i.e. a mesh size of 100) and lower and upper bounds of -5 and 18 log zoospore load, which we chose to minimize the effects of eviction on the IPM predictions (loss of individuals from the model because their predicted future loads are outside the model range, Section 4.B; Williams et al., 2012). To put these bounds in context, the log zoospore range from our experiment was (-1.14, 13.15) and the approximate log zoospore load at which *R. muscosa* begin experiencing substantial die-off in the field is at or above a log zoospore load of 9.21 (Vredenburg et al., 2010). To incorporate the discrete, uninfected stage into the IPM, we appended an extra row giving transitions of various infected individuals to an uninfected state (top-most row) and an extra column specifying the transition of uninfected individuals into various infected states (left-most column) to the 100 x 100 parasite load transition matrix described above (Merow et al., 2014*a*).

We calculated the local elasticity of the population growth rate (λ) to the lower-level regression parameters $b_{i,j}$ of the vital functions defined above by perturbing each regression parameter by $\delta = 0.001$ and calculating the elasticity as $e_{i,j} = [(\lambda_{\text{perturbed}} - \lambda_{\text{fitted}})/\delta \times b_{i,j}] \times (b_{i,j}/\lambda_{\text{fitted}})$ (Merow et al., 2014*b*). To propagate the uncertainty in our estimates of the lower-level vital rate parameters through to our estimates of the population growth rate and lower-level parameter elasticity, we took the following parametric bootstrap approach. Using standard asymptotic likelihood results (McCullagh and Nelder, 1989), we assumed that each parameter set from a vital rate function followed a multivariate normal distribution with a mean and variance-covariance matrix equal to the values given by the regression procedure used to fit the vital rate function. Next, we ran 1000 simulations in which we randomly drew the lower-level regression parameters from their respective multivariate normal distributions

and parameterized the IPM using these parameters. We then calculated either the population growth rate or the elasticity of a given lower-level parameter with these randomly drawn values and stored the result. This provided us with estimates of the population growth rate and lower-level parameter elasticity while accounting for the uncertainty in the lower-level parameters used to build the IPM. We note that this approach likely underestimates the uncertainty as it does not account for the uncertainty in the variance estimates, does not account for covariance of parameters between vital rate functions, and assumes multivariate normality. A fully Bayesian approach can capture this uncertainty more completely (Merow et al., 2014*b*; Elderd and Miller, 2015).

4.3.11 Exploring density-dependent transmission dynamics

In equation 7, we assumed density and frequency independent transmission of *Bd*. We also explored how a mass action, density-dependent transmission function affected the epizootic dynamics of *Bd*. In particular we assumed the following transmission function

$$\phi(I(x,t)) = 1 - \exp\left(-(a + \beta \int_L^U xI(x,t)dx)\right) \tag{4.13}$$

which specifies that the probability of infection at time t is dependent on the total number of zoospores present in the host population ($\int_L^U xI(x,t)dx$) at time t as well as a constant probability of infection from an environmental reservoir $\omega = 1 - \exp(-a)$. We followed the example of previous *Bd* modeling work and assumed that the *Bd* epizootic dynamics depend on the number zoospores in the aquatic environment rather than just the number of infected amphibians in a population (Briggs et al., 2010). The term $\int_L^U xI(x,t)dx$ reflects this assumption, albeit ignoring potential dynamics of free-living

zoospores. Moreover, we assumed that density dependence affects only the probability of transitioning from uninfected to infected, such that once an amphibian is infected the increase in *Bd* is independent of infected host density. This assumption is realistic if the parasite reproduction on the host swamps out the effects of reinfection from other individuals or the environment.

To explore the effects of this transmission function on epizootic dynamics, we first parameterized the density-independent portion of the IPM model using the maximum likelihood estimates of the vital rate function parameters discussed above. Because we could not estimate the density-dependent transmission function from the data we collected, we explored the effect of this function on population dynamics by choosing (ω, β) pairs on a grid and using these values to parameterize the density-dependent transmission function. The estimated values of ω used in the density-independent model suggested that ω was between 0.22 and 0.6 depending on the temperature, so we explored values of ω between 0.01 and 0.6. We did not have a good *a priori* estimate of β, so we explored β within the range 0 to 1.17 x 10^{-3}, where this upper bound was chosen arbitrarily after preliminary simulations showed that larger values of β had little effect on the population dynamics. For every (ω, β) pair, we iterated the density-dependent IPM for 120 days, which is the approximate length of the summer in the Sierra Nevada during which *Bd* epizootics tend to occur (Briggs et al., 2005, 2010). We initialized each population with 100 uninfected individuals and for each combination of (ω, β) we calculated the proportion of surviving amphibians and the prevalence of *Bd* infection at the end of the epizootic.

4.4 Results

4.4.1 Vital rate functions

Increasing log zoospore load x significantly decreased survival probability of amphibians (Fig. 4.2A; $\chi^2_{df=1} = 12.197$, $p = 0.0005$; Table 1).

Both temperature and log zoospore load at time t significantly increased log zoospore load at time $t+1$ (Likelihood Ratio Test (LRT) for load at time t: $\chi^2_{df=1} = 196.36$, $p < 0.0001$; LRT for temperature: $\chi^2_{df=1} = 13.56$, $p = 0.0002$). Moreover, log zoospore load at time t was important for describing the variance structure of the growth function, as compared to a model with constant variance structure (LRT comparing full model to model with constant variance: $\chi^2_{df=1} = 9.8$, $p = 0.0017$; Table 1; Fig. 4.2B, Fig. 4.3).

Temperature and log zoospore load were both highly significant predictors of whether an amphibian would clear *Bd* infection in a given time step (temperature: $\chi^2_{df=1} = 14.555$, $p = 0.0001$; log zoospore load: $\chi^2_{df=1} = 23.701$, $p < 0.0001$; Fig. 4.3). Amphibians were more likely to clear infection at lower temperatures and when the load at time t was smaller.

Increasing temperature significantly increased the mean and variance of the initial infection load distribution $G_0(x')$ (temperature effect on mean: $t_{df=41} = 2.53$, $p = 0.015$; temperature effect on variance: LRT comparing model with variance structure to without: $\chi^2_{df=1} = 6.00$, $p = 0.0143$; Fig. 4.3).

Finally, increasing temperature significantly increased the probability of infection ϕ ($\chi^2_{df=1} = 6.0361$, $p = 0.014$; Table 1).

4.4.2 Laboratory dynamics of amphibians and *Bd*

The parameterized IPM model predicted that individual amphibians at low temperatures would survive significantly longer than amphibians at high temperatures, with the largest difference being when log zoospore loads were low (Fig. 4.4A). Over a summer epizootic, amphibian populations at low temperatures experienced a minimal effect of *Bd*-induced population declines ($\lambda \approx 1$), while amphibians at higher temperatures experience substantially more rapid declines, with large uncertainty around these estimates (Fig. 4.4B). Elasticity analysis on the lower-level parameters used in the vital rate functions showed that overall population growth rate was most sensitive to proportional changes in the growth rate of *Bd* (the parameters of the growth function; Fig. 4.S5) as well as the pathogenicity of *Bd* and the threshold at which *Bd*-induced mortality began to occur (the parameters of the survival function; Fig. 4.S5).

The IPM model also allowed us to examine how the stable log zoospore distribution of *Bd* on surviving hosts changed with temperature (Fig. 4.5). For surviving, infected amphibians, the mean infection intensity increased with temperature, but the variance to mean ratio decreased with temperature (Fig. 4.5), consistent with experimental and model results showing that hosts experienced greater *Bd*-induced mortality at higher temperatures. This is also consistent with previous theoretical results from macroparasite models which predict that increased parasite-induced host mortality generally decreases the variance to mean ratio (Barbour and Pugliese, 2000).

4.4.3 Effects of density-dependent transmission on epizootic dynamics

The effect of density-dependent transmission on *Bd*-*R. muscosa* population dynamics varied with temperature, the probability of infection from the environmental reservoir (ω) and the transmission coefficient (β). In general, over the range of density-dependent transmission values we examined, density-dependent transmission had little effect on prevalence and the proportion of population decline over the course of a summer epizootic (Fig. 4.S7-S8). In contrast, the probability of infection from the environment had a large effect on both prevalence patterns and population decline (Fig. 4.S7-S8). Given a probability of infection from the environment above approximately 0.15, increasing density-dependent transmission had very little effect on *Bd* prevalence or *R. muscosa* population decline. Over the parameter space we examined, the density-dependent transmission model predicted that populations at 12 °C will experience a maximum of a 20% population decline over the course of an epizootic with 70% prevalence, while populations at 20 °C will experience a greater than 80% population decline with close to 100% prevalence (Fig. 4.S7-S8).

4.5 Discussion

Integral projection models provide an ideal framework to model diseases that do not fall neatly into the microparasite/macroparasite dichotomy. By taking an intermediate approach between individual-based disease models which explicitly track the parasite load on every individual in a population (Briggs et al., 2010) and classic macroparasite/microparasite models which only track the total number of hosts and parasites in a

population (Anderson and May, 1978), IPMs can elegantly investigate population outcomes of infectious diseases while still incorporating critical information about disease dynamics at the individual-level (Metcalf et al., 2015). While the IPM approach can theoretically be used to explore the dynamics in any macroparasite or microparasite system, we believe it will be practically most useful in host-parasite systems where the growth rate of a parasite is slow enough that measurements of parasite load at time t and $t+1$ are on the same time scale as the growth rate of the parasite. This allows for empirical estimation of the vital rate functions and an investigation regarding how these vital rate functions vary with environmental factors such as temperature and/or differ between host populations in which a disease is established or invading.

We used the host-parasite IPM model to explore the consequences of different temperatures on *R. muscosa*-*Bd* dynamics over the course of an epizootic. The effect of temperature on *Bd* growth is well-known both in culture and on amphibian hosts (Longcore et al., 1999; Piotrowski et al., 2004; Berger et al., 2004; Andre et al., 2008; Raffel et al., 2012) and previous work has estimated the expected time to death of amphibians infected with *Bd* over various different temperatures (Berger et al., 2004; Andre et al., 2008). However, the effect of temperature-*Bd* interactions on amphibians at the population level is much less clear (Rohr and Raffel, 2010; Knapp et al., 2011). Using an IPM model, we were able to make specific, quantitative predictions about how temperature and transmission dynamics affected population growth rates of *Rana muscosa*.

The density-independent IPM model predicted that population-level growth rate decreased with increasing temperature and naive populations at or above about 18 degrees had a 50% chance of experiencing an 80% decline or greater over the course of a summer epizootic. This result likely represents a best case scenario for *Rana muscosa* as this density-independent model does not account for *Bd* transmission dynamics

(Rachowicz and Briggs, 2007) or additional factors leading to increased frog mortality or *Bd*-susceptibility in the field. Our elasticity analysis showed that the population-level growth rate was most sensitive to proportional changes in parameters relating to the *Bd* growth function and the survival function. If *in situ* factors slightly reduced the *Bd*-load at which frogs began experiencing disease-induced mortality, for example, *R. muscosa* populations could experience extirpation during a summer epizootic for a wide range of temperatures, which would be consistent with the patterns observed in the field (Knapp et al., 2011). In particular, we assumed a temperature-independent survival function in the IPM model (described in *Vital rate functions*) and including temperature-dependence into this function would have significant impacts the ability of *R. muscosa* populations to persist through an epizootic.

We extended this density-independent IPM to explore how density-dependent transmission and transmission from an environmental reservoir affected population dynamics. Our results suggest that density-dependent transmission had a small effect on the population dynamics of *Bd* epizootics, particularly when an environmental reservoir was present. While this result is largely due to our assumption that density-dependent transmission does not effect the the growth of *Bd* on an already infected frog, it is consistent with predictions from a fully individual-based model that predicts that density manipulations (i.e. culling infected frogs) will likely have little effect on mitigating population outcomes during *Bd* epizootics in this system (Drawert et al., 2015). A natural next step will be to use this IPM to investigate how varying temperature regimes and *R. muscosa* demography affect the persistence of *R. muscosa* populations infected with *Bd* over longer time scales. In general, the question of how temperature interacts with *Bd* and in turn affects amphibian host persistence is a critical question in amphibian conservation (Rohr and Raffel, 2010) and IPMs provide a novel means by which this

question can be quantitatively addressed.

In addition to these population-level predictions, host-parasite IPMs also allow for explicit predictions about how the distribution of parasites loads over hosts changes with different vital parameters and/or over the course of an epizootic or enzootic. Macroparasite models have long recognized the importance of the distribution of parasite loads over hosts for determining the dynamics of host-parasite interactions (Anderson and May, 1978; Tompkins et al., 2002), and classic macroparasite models addressed this by using a statistical distribution (often negative binomial, Shaw et al., 1998) and then looking at how different levels of parasite aggregation affected host-parasite dynamics (Anderson and May, 1978; Kretzschmar and Alder, 1993). These approaches have been extended to include fluctuating aggregation (Rosà and Pugliese, 2002; Rosà et al., 2003), but still rely on explicitly defining the shape of the host-parasite distribution. In contrast, IPMs do not assume a host-parasite distribution, rather one emerges as a result of the vital functions specified when parameterizing the model. Therefore, one can explore how sensitive the aggregation of the host-parasite distribution is to different vital function parameters, providing an intriguing way to parse the contribution of different processes to parasite aggregation. Moreover, as it is straightforward to include seasonal fluctuations and/or environmental stochasticity into the IPM framework (Rees and Ellner, 2009; Eager et al., 2013), more complex predictions of aggregation patterns, such as the fluctuation of parasite aggregation over time (Scott, 1987; Rosà and Pugliese, 2002), could be explored.

Using the parameterized IPM for *Bd*-*R. muscosa* we examined how the distribution of *Bd*-loads changed with temperature. The IPM showed that fundamental insight from macroparasite distributions also applies to *Bd*. For example, as predicted by macroparasite models (Barbour and Pugliese, 2000), increasing *Bd*-induced host mortality with

increasing temperature decreased the aggregation of *Bd* across hosts and reduced positive skew as individuals with high *Bd* loads were removed from the population through mortality. In fact, a sensitivity analysis of the variance to mean ratio of the *Bd*-load distributions showed that this measure of aggregation became progressively more sensitive to the survival function as temperature increased and more frogs experienced *Bd* load-dependent mortality (Fig. 4.S6). In addition, the variance to mean ratio was more sensitive to the variance in the growth function ($\nu_{0,1}$ and $c_{0,1}$) than the variance in the initial infection burden function ($\nu_{0,3}$ and $c_{0,3}$, Fig. 4.S6), suggesting that explaining the individual-level heterogeneity in *Bd* growth rate may be more important for understanding the shape of the *Bd*-load distribution than explaining the individual-level heterogeneity in the load of *Bd* at initial infection. The IPM approach highlights the importance of this unexplained variance in the *Bd* growth function and future studies could identify whether this heterogeneity is due to biological factors such as differences in immune responses among hosts or methodological factors such as quantitative PCR error when measuring *Bd* load.

In addition to allowing for a more rigorous analysis of parasite aggregation, an IPM approach can be used to examine a variety of different classic patterns in host-parasite systems. For example, host age can easily be included as an additional host attribute (Childs et al., 2003, 2004), such that IPMs could then be used to examine observed patterns between parasite intensity and host age (i.e. age-intensity profiles, Duerr et al., 2003). Similarly, host-heterogeneity in susceptibility could be included as an additional host attribute such that IPMs could be used to explore non-linear dose-response relationships (Dwyer et al., 1997; Gomes et al., 2014). We also discuss in Section 4.C how R_0, a canonical epidemiological measure of the ability of a parasite to invade a fully susceptible host population (Diekmann et al., 1990), can be calculated from the host-

parasite IPM. While these are just a few examples, the theoretical application of IPMs for exploring observed host-parasite patterns is extensive.

While this study focused on using IPMs to describe epizootic dynamics of amphibian chytrid fungus, there are a variety of other wildlife diseases in which host-parasite IPMs could be applicable to explore the population and evolutionary outcomes of infection. For example, Tasmanian devils *Sarcophilus harrisii* are threatened with extinction by an infectious cancer, Tasmanian devil facial tumor disease (McCallum et al., 2009). A critical question for management is to predict the impact of the disease as it enters currently uninfected populations and to investigate evidence of selection for increased resistance to infection or reduced tumor growth rates. Intensive mark-recapture data are available, enabling the estimation of survival rates of infected and uninfected animals, together with transition rates from uninfected to infected states (Hamede et al., 2012). In addition, measurements of tumor size are taken from all infected animals at every capture opportunity and repeated tumor measurements are available for a substantial number of individuals, which could be used to estimate the tumor growth function. One could examine whether the death rate of infected devils is related to the size of the tumor, and then use the IPMs to examine how differences in tumor growth among populations or over time might alter the dynamics of devil populations. It is highly likely that the death rate of infected devils is related to the size of the tumor. This problem may therefore be well-suited for an IPM approach, permitting more accurate modeling of the impact of the tumor on devil population dynamics.

Similarly, an IPM approach could also be taken to explore various aspects of the ecology and evolution of bats affected by white-nose syndrome, an emerging fungal disease of North American bats (Blehert et al., 2008). White-nose syndrome is characterized by intense transmission, such that nearly 100% of bats of multiple species may become

infected during the first winter after the fungus reaches a site (Langwig et al., 2015*b*). Mortality, which occurs 70-100 days after initial infection in lab studies (Warnecke et al., 2012), usually occurs in mid to late winter when fungal loads are highest (Langwig et al., 2015*a*). IPMs could be fit to pathogen loads and population dynamics of bats to explore how temperature and humidity influence pathogen growth and disease impacts (Langwig et al., 2012). Through modification of the growth function and survival function, IPMs could be used to determine whether persistence of some stabilizing populations could be explained by resistance or tolerance, or other factors affecting host-parasite interactions.

In conclusion, the use of IPMs can answer important questions regarding host-pathogen interactions in wildlife and plant disease. Moreover, IPMs can provide new insight into many classic micro and macroparasite patterns such as the distribution of parasites across hosts, age-intensity profiles, and the dynamics of infection prevalence. By bridging the gap between micro and macroparasites, IPMs provide an exciting new frontier in modeling wildlife disease.

4.6 Acknowledgments

This work was supported by the National Science Foundation (DEB-1336290, DEB-1115895, and and DEB-1316549), and the National Institutes of Health (R01GM109499). These ideas were developed during an NCEAS working group Fungal pathogens and disease-induced extinction: Are fungal diseases different?. MW was supported by the National Science Foundation Graduate Research Fellowship (Grant No.DGE 1144085).

References

Allen, L. J., and P. van den Driessche. 2008. The basic reproduction number in some discrete-time epidemic models. Journal of Difference Equations and Applications **14**:1127–1147.

Anderson, R. M., and R. M. May. 1978. Regulation and stability of host-parasite interactions: I. Regulatory processes. Journal of Animal Ecology **47**:219–247.

Anderson, R. M., and R. M. May. 1979. Population biology of infectious diseases: Part I. Nature **280**:361 – 367.

Andre, S. E., J. Parker, and C. J. Briggs. 2008. Effect of temperature on host response to Batrachochytrium dendrobatidis infection in the mountain yellow-legged frog (Rana muscosa). Journal of Wildlife Diseases **44**:716–720.

Barbour, A. D., and A. Pugliese. 2000. On the variance-to-mean ratio in models of parasite distributions. Advances in Applied Probability **32**:701–719.

Berger, L., R. Speare, H. B. Hines, G. Marantelli, A. D. Hyatt, K. R. McDonald, L. F. Skerratt, V. Olsen, J. M. Clarke, G. Gillespie, M. Mahony, N. Sheppard, C. Williams, and M. J. Tyler. 2004. Effect of season and temperature on mortality in amphibians due to chytridiomycosis. Australian Veterinary Journal **82**:434–9.

Blehert, D. S., A. C. Hicks, M. Behr, C. U. Meteyer, B. M. Berlowski-zier, E. L. Buckles, J. T. H. Coleman, S. R. Darling, A. Gargas, R. Niver, J. C. Okoniewski, R. J. Rudd, and B. Ward. 2008. Bat White-Nose Syndrome : An Emerging Fungal Pathogen ? Science **323**:227.

Boyle, D. G., D. B. Boyle, V. Olsen, J. A. T. Morgan, and A. D. Hyatt. 2004. Rapid quantitative detection of chytridiomycosis (Batrachochytrium dendrobatidis) in amphibian samples using real-time Taqman PCR assay. Diseases of Aquatic Organisms **60**:141–8.

Briggs, C. J., R. A. Knapp, and V. T. Vredenburg. 2010. Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. Proceedings of the National Academy of Sciences of the United States of America **107**:9695–700.

Briggs, C. J., V. T. Vredenburg, R. A. Knapp, and L. J. Rachowicz. 2005. Investigating the population-level effects of chytridiomycosis: an emerging infectious disease of amphibians. Ecology **86**:3149–3159.

Bruno, J. F., S. P. Ellner, I. Vu, K. Kim, and C. D. Harvell. 2011. Impacts of aspergillosis on sea fan coral demography : modeling a moving target. Ecological Monographs **81**:123–139.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. Second edition. Sinauer, Sunderland, MA.

Childs, D. Z., M. Rees, K. E. Rose, P. J. Grubb, and S. P. Ellner. 2003. Evolution of complex flowering strategies: an age- and size-structured integral projection model. Proceedings of the Royal Society B: Biological Sciences **270**:1829–1838.

Childs, D. Z., M. Rees, K. E. Rose, P. J. Grubb, and S. P. Ellner. 2004. Evolution of size-dependent flowering in a variable environment: construction and analysis of a stochastic integral projection model. Proceedings of the Royal Society B: Biological Sciences **271**:425–34.

Cooch, E. G., P. B. Conn, S. P. Ellner, A. P. Dobson, and K. H. Pollock. 2012. Disease dynamics in wild populations: modeling and estimation: a review. Journal of Ornithology **152**:485–509.

Coulson, T. 2012. Integral projections models, their construction and use in posing hypotheses in ecology. Oikos **121**:1337–1350.

Dahlgren, J. P., M. B. García, and J. Ehrlén. 2011. Nonlinear relationships between vital rates and state variables in demographic models. Ecology **92**:1181–1187.

Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz. 1990. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology **28**:365–382.

Drawert, B., M. Griesemer, L. Petzold, and C. J. Briggs. 2015. Using stochastic epidemiological models to evaluate strategies to save endandered amphibians. Proceedings of the National Academy of Sciences **In Review**.

Duerr, H. P., K. Dietz, and M. Eichner. 2003. On the interpretation of ageintensity profiles and dispersion patterns in parasitological surveys. Parasitology **126**:87–101.

Dwyer, G., J. Elkinton, and J. Buonaccorsi. 1997. Host Heterogeneity in Susceptibility and Disease Dynamics: Tests of a Mathematical Model. The American Naturalist **150**:685–707.

Eager, E. A., C. V. Haridas, D. Pilson, R. Rebarber, and B. Tenhumberg. 2013. Disturbance frequency and vertical distribution of seeds affect long-term population dynamics: a mechanistic seed bank model. The American naturalist **182**:180–90.

Easterling, M. R., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new structured population model. Ecology **81**:694–708.

Elderd, B. D., and T. E. X. Miller. 2015. Quantifying demographic uncertainty : Bayesian methods for integral projection models. Ecology **86**:125–144.

Fisher, M. C., D. A. Henk, C. J. Briggs, J. S. Brownstein, L. C. Madoff, S. L. McCraw, and S. J. Gurr. 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature **484**:186–94.

Gomes, M. G. M., M. Lipsitch, A. R. Wargo, G. Kurath, C. Rebelo, G. F. Medley, and A. Coutinho. 2014. A missing dimension in measures of vaccination impacts. PLoS Pathogens **10**:e1003849.

Hamede, R., S. Lachish, K. Belov, G. Woods, A. Kreiss, A. M. Pearse, B. Lazenby, M. Jones, and H. Mccallum. 2012. Reduced Effect of Tasmanian Devil Facial Tumor Disease at the Disease Front. Conservation Biology **26**:124–134.

Isham, V. 1995. Stochastic models of host-macroparasite interaction. The Annals of Applied Probability **5**:720–740.

Kilpatrick, A. M., C. J. Briggs, and P. Daszak. 2010. The ecology and impact of chytridiomycosis: an emerging disease of amphibians. Trends in Ecology and Evolution **25**:109–118.

Klepac, P., and H. Caswell. 2011. The stage-structured epidemic: Linking disease and demography with a multi-state matrix approach model. Theoretical Ecology **4**:301–319.

Knapp, R. A., C. J. Briggs, T. C. Smith, and J. R. Maurer. 2011. Nowhere to hide: impact of a temperature-sensitive amphibian pathogen along an elevation gradient in the temperate zone. Ecosphere **2**:art93.

Kretzschmar, M., and F. R. Alder. 1993. Aggregated distributions in models for patchy populations. Theoretical Population Biology **43**:1–30.

Langwig, K. E., W. F. Frick, J. T. Bried, A. C. Hicks, T. H. Kunz, and A. M. Kilpatrick. 2012. Sociality, density-dependence and microclimates determine the persistence of populations suffering from a novel fungal disease, white-nose syndrome. Ecology Letters **15**:1050–1057.

Langwig, K. E., W. F. Frick, R. Reynolds, K. L. Parise, K. P. Drees, J. R. Hoyt, T. L. Cheng, T. H. Kunz, J. T. Foster, and A. M. Kilpatrick. 2015*a*. Host and pathogen ecology drive the seasonal dynamics of a fungal diseas , white-nose syndrome. Proceedings of the Royal Society B: Biological Sciences **282**:20142335.

Langwig, K. E., J. R. Hoyt, K. L. Parise, J. Kath, D. Kirk, W. F. Frick, J. T. Foster, and A. M. Kilpatrick. 2015*b*. Invasion dynamics of white-nose syndrome fungus, midwestern United States, 20122014. Emerging Infectious Diseases **21**:1023–1026.

Longcore, J. E., A. P. Pessier, D. K. Nichols, and J. E. Longcorel. 1999. Batrachochytrium dendrobatidis gen. et sp . nov., a chytrid pathogenic to amphibians. Mycologia **91**:219–227.

May, R. M., and R. M. Anderson. 1978. Regulation and stability of host-parasite population interactions: II. Destabilizing processes. Journal of Animal Ecology **47**:249–267.

May, R. M., and R. M. Anderson. 1979. Population biology of infectious disease: Part II. Nature **280**:455 – 461.

McCallum, H., N. Barlow, and J. Hone. 2001. How should pathogen transmission be modelled? Trends in Ecology and Evolution **16**:295–300.

McCallum, H., M. Jones, C. Hawkins, R. Hamede, S. Lachish, D. L. Sinn, N. Beeton, and B. Lazenby. 2009. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology **90**:3379–3392.

McCallum, H. I., 2000. Host-pathogen and host-parasite models. Chapter chapter 10, pages 284–312 *in* J. H. Lawton and G. E. Likens, editors. Population Parameters: Estimation for Ecological Models. Blackwell Science Ltd.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models. Second edition. Chapman & Hall, New York.

Merow, C., J. P. Dahlgren, C. J. E. Metcalf, D. Z. Childs, M. E. K. Evans, E. Jongejans, S. Record, M. Rees, R. Salguero-Gómez, and S. M. Mcmahon. 2014*a*. Advancing population ecology with integral projection models: A practical guide. Methods in Ecology and Evolution **5**:99–110.

Merow, C., A. M. Latimer, A. M. Wilson, S. M. McMahon, A. G. Rebelo, and J. A. Silander. 2014*b*. On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. Ecography **37**:1167–1183.

Metcalf, C. J. E., A. L. Graham, M. Martinez-Bakker, and D. Z. Childs. 2015. Op-

portunities and challenges of Integral Projection Models for modelling host-parasite dynamics. Journal of Animal Ecology **85**:343–355.

Metcalf, C. J. E., S. M. McMahon, R. Salguero-Gómez, and E. Jongejans. 2013. IPMpack : an R package for integral projection models. Methods in Ecology and Evolution **4**:195–200.

Oli, M. K., M. Venkataraman, P. A. Klein, L. D. Wendland, and M. B. Brown. 2006. Population dynamics of infectious diseases: A discrete time model. Ecological Modelling **198**:183–194.

Piotrowski, J. S., S. L. Annis, and J. E. Longcore. 2004. Physiology of Batrachochytrium dendrobatidis, a chytrid pathogen of amphibians. Mycologia **96**:9–15.

Pugliese, A., R. Rosà, and M. L. Damaggio. 1998. Analysis of model for macroparasitic infection with variable aggregation and clumped infections. Journal of Mathematical Biology **36**:419–47.

Rachowicz, L. J., and C. J. Briggs. 2007. Quantifying the disease transmission function: effects of density on Batrachochytrium dendrobatidis transmission in the mountain yellow-legged frog Rana muscosa. The Journal of Animal Ecology **76**:711–21.

Raffel, T. R., J. M. Romansic, N. T. Halstead, T. A. McMahon, M. D. Venesky, and J. R. Rohr. 2012. Disease and thermal acclimation in a more variable and unpredictable climate. Nature Climate Change **3**:146–151.

Rees, M., D. Z. Childs, and S. P. Ellner. 2014. Building integral projection models: A user's guide. Journal of Animal Ecology pages 528–545.

Rees, M., and S. P. Ellner. 2009. Integral projection models for populations in temporally varying environments. Ecological Monographs **79**:575–594.

Roberts, M. G., G. Smith, and B. T. Grenfell, 1995. Mathematical models for macroparasites of wildlife. Pages 177–208 *in* B. T. Grenfell and A. P. Dobson, editors. Ecology of Infectious Diseases in Natural Populations. Cambridge University Press, Cambridge, United Kingdom.

Rohani, P., R. Breban, D. E. Stallknecht, and J. M. Drake. 2009. Environmental transmission of low pathogenicity avian influenza viruses and its implications for pathogen invasion. Proceedings of the National Academy of Sciences of the United States of America **106**:10365–10369.

Rohr, J. R., and T. R. Raffel. 2010. Linking global climate and temperature variability to widespread amphibian declines putatively caused by disease. Proceedings of the National Academy of Sciences of the United States of America **107**:8269–8274.

Rollins-Smith, L. A. 2009. The role of amphibian antimicrobial peptides in protection of amphibians from pathogens linked to global amphibian declines. Biochimica et Biophysica Acta **1788**:1593–9.

Rosà, R., and A. Pugliese. 2002. Aggregation, stability, and oscillations in different models for host-macroparasite interactions. Theoretical Population Biology **61**:319–34.

Rosà, R., A. Pugliese, A. Villani, and A. Rizzoli. 2003. Individual-based vs. deterministic models for macroparasites: host cycles and extinction. Theoretical Population Biology **63**:295–307.

Scott, M. E. 1987. Temporal changes in aggregation: a laboratory study. Parasitology **94 (Pt 3)**:583–595.

Shaw, D. J., B. T. Grenfell, and A. P. Dobson. 1998. Patterns of macroparasite aggregation in wildlife host populations. Parasitology **117**:597–610.

Skerratt, L. F., L. Berger, R. Speare, S. Cashins, K. R. McDonald, A. D. Phillott, H. B. Hines, and N. Kenyon. 2007. Spread of Chytridiomycosis Has Caused the Rapid Global Decline and Extinction of Frogs. EcoHealth **4**:125–134.

Smith, M. J., S. Telfer, E. R. Kallio, S. Burthe, A. R. Cook, X. Lambin, and M. Begon. 2009. Host-pathogen time series data in wildlife support a transmission function between density and frequency dependence. Proceedings of the National Academy of Sciences of the United States of America **106**:7905–9.

Tompkins, D. M., A. P. Dobson, P. Arneberg, M. Begon, I. M. Cattadori, J. V. Greenman, J. A. P. Heesterbeek, P. J. Hudson, D. Newborn, A. Pugliese, A. P. Rizzoli, R. Rosa, F. Rosso, and K. Wilson, 2002. Parasites and host population dynamics. Chapter 3, pages 45–62 *in* P. J. Hudson, A. Rizzoli, B. T. Grenfell, H. Heessterbeck, and A. P. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Voyles, J., L. Berger, S. Young, R. Speare, R. Webb, J. Warner, D. Rudd, R. Campbell, and L. Skerratt. 2007. Electrolyte depletion and osmotic imbalance in amphibians with chytridiomycosis. Diseases of Aquatic Organisms **77**:113–118.

Voyles, J., S. Young, L. Berger, C. Campbell, W. F. Voyles, and A. Dinudom. 2009. Pathogenesis of Chytridiomycosis a cause of catastrophic amphibian declines. Science **326**:5–8.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs. 2010. Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proceedings of the National Academy of Sciences of the United States of America **107**:9689–94.

Warnecke, L., J. M. Turner, T. K. Bollinger, J. M. Lorch, V. Misra, P. M. Cryan, G. Wibbelt, D. S. Blehert, and C. K. R. Willis. 2012. Inoculation of bats with European Geomyces destructans supports the novel pathogen hypothesis for the origin of white-nose syndrome. Proceedings of the National Academy of Sciences **109**:6999–7003.

Weatherly, N. F. 1971. Effects on litter size and litter survival in Swiss mice infected with Trichinella spiralis during gestation. The Journal of Parasitology **57**:298–301.

Wilber, M. Q., S. B. Weinstein, and C. J. Briggs. 2016. Detecting and quantifying parasite-induced host mortality from intensity data : method comparisons and limitations. International Journal for Parasitology **46**:59–66.

Williams, J. T., T. E. X. Miller, and S. P. Ellner. 2012. Avoiding unintentional eviction from integral projection models. Ecology **93**:2008–2014.

Woodhams, D. C., R. A. Alford, C. J. Briggs, M. Johnson, and L. A. Rollins-Smith. 2008. Life-history trade-offs influence disease in changing climates: Strategies of an amphibian pathogen. Ecology **89**:1627–1639.

Table 4.1: Vital rate parameters used to parameterize the density-independent integral projection model. *logit* specifies a logistic link, x is log zoospore load, and T is temperature.

Description	**Functional Form**	**Parameters**	**Details of Parameterization**
Infected survival function, $s(x)$	$\text{logit}[s(x)] = b_{0,0} + b_{1,0}x$	$b_{0,0} = 11.824$ $b_{1,0} = -0.8605$	Logistic Regression
Uninfected survival probability, s_0	Constant	$s_0 = 1$	Briggs *et al.* 2005
Growth function, $G(x', x)$	$\mu(x,T) = b_{0,1} + b_{1,1}x + b_{2,1}T$ $\sigma^2(x) = \nu_{0,1}\exp(2c_{0,1}x)$	$b_{0,1} = 0.012$ $b_{1,1} = 0.799$ $b_{2,1} = 0.092$ $\nu_{0,1} = 5.92$ $c_{0,1} = -0.049$	Generalized Least Squares
Loss of infection function, $l(x)$	$\text{logit}[l(x,T)] = b_{0,2} + b_{1,2}x + b_{2,2}T$	$b_{0,2} = 1.213$ $b_{1,2} = -0.472$ $b_{2,2} = -0.151$	Logistic Regression
Initial infection burden function, $G_0(x')$	$\mu(T) = b_{0,3} + b_{1,3}T$ $\sigma^2(T) = \nu_{0,3}\exp(2c_{0,3}T)$	$b_{0,3} = 0.642$ $b_{1,3} = 0.137$ $\nu_{0,3} = 0.59$ $c_{0,3} = 0.063$	Generalized Least Squares
Transmission function, ϕ	$\text{logit}[\phi(T)] = b_{0,4} + b_{1,4}T$	$b_{0,4} = -1.66$ $b_{1,4} = 0.102$	Logistic Regression

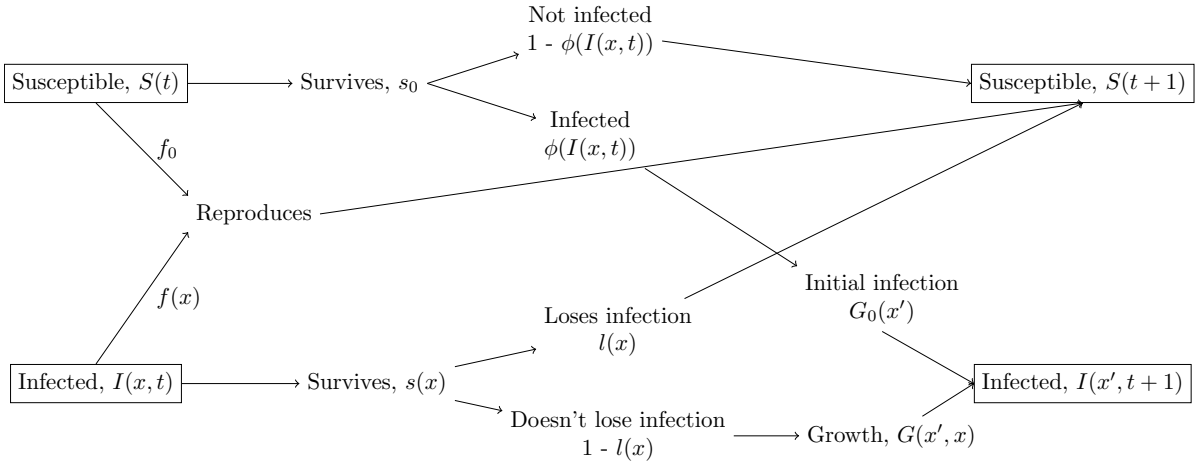


Figure 4.1: Life history flow chart for the host-parasite Integral Projection Model. The chart shows how an infected host with a parasite load of x at time t can transition to an infected host or susceptible/uninfected host with a parasite load of x' or 0, respectively, at time $t+1$. The chart also shows how an susceptible/uninfected host at time t can transition to an infected host or susceptible/uninfected host with a parasite load of x' or 0, respectively, at time $t+1$.

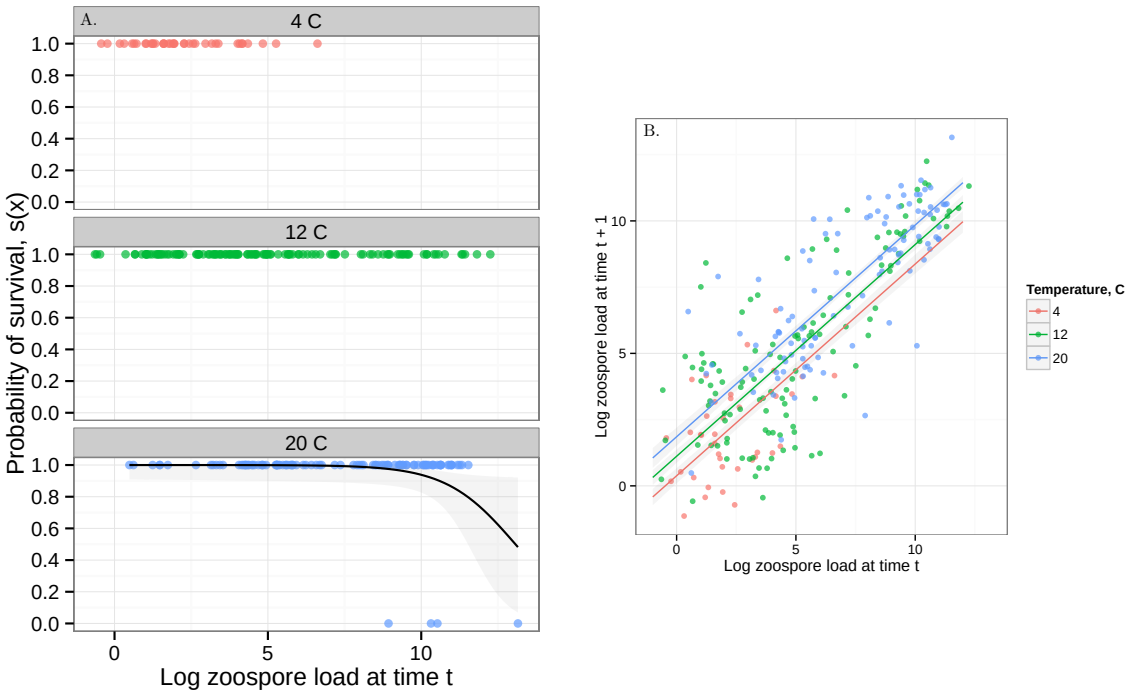


Figure 4.2: **A.** The laboratory data used to estimate the survival function $s(x)$. Each panel gives a different temperature and each point gives the load of an individual frog at time t and whether it survived to time $t+1$. A value of 1 indicates that a frog survived and a value of 0 indicates that it died. No frogs died in temperature treatments 4 and 12 °C. The black line in the 20 °C plot gives the fit of the temperature-independent survival function used in the analysis, plus or minus the standard error about the prediction. **B.** The laboratory *Bd* growth data and corresponding temperature-dependent growth function $G(x', x)$ from the *Bd-Rana muscosa* laboratory experiment. Each point gives the log zoospore load on an individual at time t and time $t+1$. The different colors show different temperatures. The corresponding lines give the predicted growth function for a given temperature along with the standard error about the predicted mean. Growth of *Bd* on an individual frog increases with both temperature and the number of zoospores at time t. Alternative models for this growth function are discussed in Section 4.A.

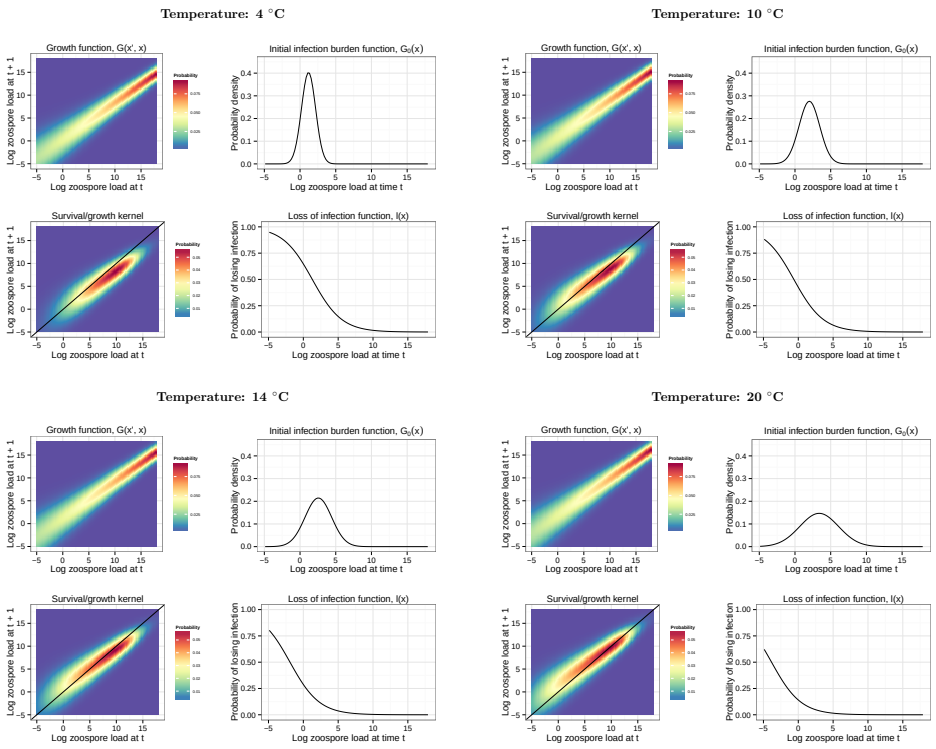


Figure 4.3: The growth function $G(x', x)$, loss of infection function $l(x)$, initial infection burden function $G_0(x')$ and the survival/growth kernel $(G(x', x)s(x))$ used to parameterize the *Bd-Rana muscosa* Integral Projection Model for temperatures between 4 and 20 °C. The four temperatures shown were chosen to illustrate how the various vital rate functions change with temperature. Because each vital rate function shown is a linear function of temperature (see *Vital rate functions*) we were not restricted to choosing the 3 temperatures used to fit the vital rate functions (4, 12, and 20 °C) and could chose any temperature between 4 and 20 °C. The black line on the survival/growth kernel plots is a one to one line representing stasis: above this line the *Bd* load on a host gets larger in a time step and below this line the *Bd* load on a host gets smaller in a time step.

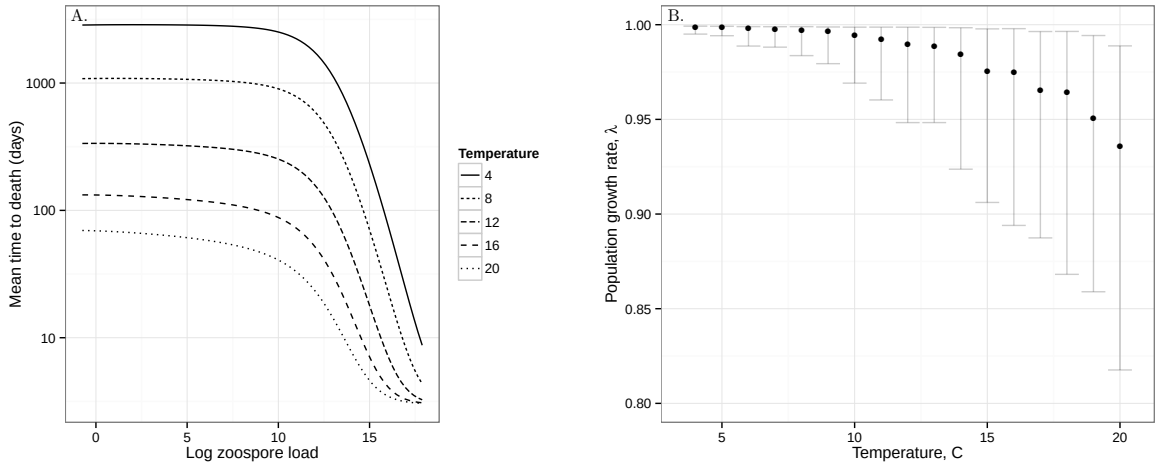


Figure 4.4: The host-parasite Integral Projection Model predictions for **A.** how the expected time to *Bd*-induced *Rana muscosa* death varies with log zoospore load and temperature and **B.** how the population growth rate (λ) of *R. muscosa* varies with temperature. The black points are the median population growth rate for 1000 simulations that account for the uncertainty in the vital rate function parameters. The error bars give the first and third quartiles of λ from these simulations.

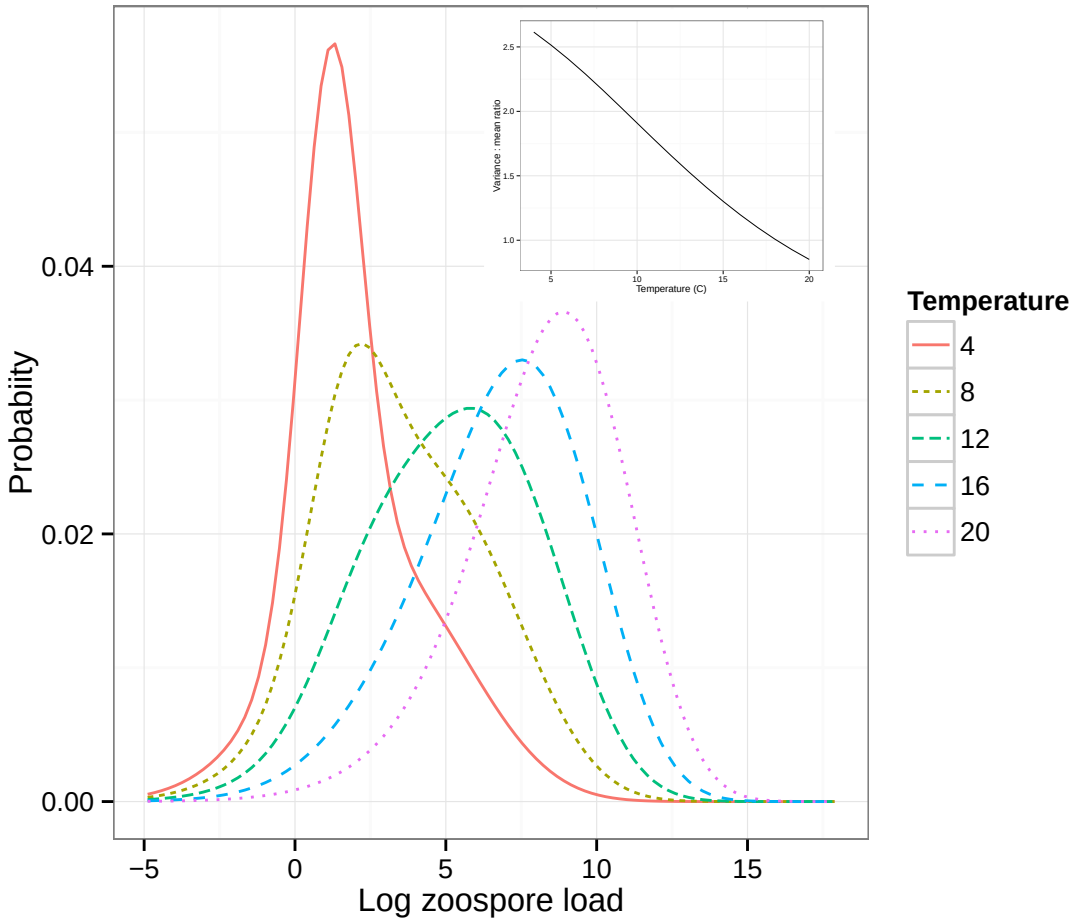


Figure 4.5: The stable *Bd* load distribution for infected *Rana muscosa* as predicted by the parameterized host-parasite Integral Projection Model for various different temperatures. The inset plot shows that the variance to mean ratio of this distribution decreases with temperature.

4.A Model selection for vital rate functions

4.A.1 Survival function: $s(x)$

In the main text, we parameterized the survival function $s(x)$ using only individuals at 20 °C. As discussed in the main text, we chose to do this because individuals at this temperature were the only frogs that experienced mortality and we have substantial alternative evidence that the load-survival relationship between *R. muscosa* and *Bd* is not strongly temperature dependent. An alternative way that we could have parameterized the model was using all of the temperature data (4, 12, and 20 °C), but without including an effect of temperature. Given the link function $\text{logit}[s(x)] = b_0 + b_1 x$, the parameters change from $b_{0,\text{only 20 °C}} = 11.5973$ (SE: 4.74) to $b_{0,\text{all temperatures}} = 11.8241$ (SE; 4.16) and $b_{1,\text{only 20 °C}} = -0.8873$ (SE: 0.45) to $b_{1,\text{all temperatures}} = -0.8605$ (SE: 0.39) (Figure 4.S1). Despite these seemingly small differences, our elasticity analysis shows that small changes in this survival function can have large effects on the ability of *R. muscosa* to persist through an epizootic.

4.A.2 Growth Function: $G(x', x)$

We explored a variety of different models for the growth functions $G(x', x)$ (Table 4.S1, Figure 4.S2). We did identify a more complex model than the model described in the manuscript that included a quadratic term for log zoospore size and an interaction between temperature and log zoospore load (Table 4.S1; Model 6). We chose to use the linear model (Model 2) because 1) the quadratic model was highly specific for the data used to fit the model and did not give a generalizable *Bd* growth curve (e.g. exponential growth) and 2) for a given temperature the quadratic model did not allow for realistic extrapolation beyond the range of the data used to fit the model because for small log-zoospore loads the function predicts that increasing the log zoospore load at time t decreases log zoospore load at time $t + 1$ (i.e. when you are on the decreasing arm of the quadratic function, Figure 4.S4). This is not a biologically reasonable pattern.

Despite these drawbacks, we ran the IPM analysis described in the paper using this quadratic growth model. We accounted for the zero derivative of the quadratic function by defining the growth function as the following piecewise function

$$G(x', x) \text{ if } \frac{d\mu(x,T)}{dx} > 0 \tag{4.14}$$

$$G(x', x_0) \text{ if } \frac{d\mu(x,T)}{dx} \leq 0 \tag{4.15}$$

where x_0 is the log zoospore load at which the derivate of $\mu(x,T) = b_0 + b_1 x + b_2 x^2 + b_3 T + b_4 xT$ is equal to 0. Analyzing the IPM with this growth function in place of the growth function used in the main text provided qualitatively similar results: population growth rate decreased with increasing temperature and the population growth rate was most sensitive to proportional changes in the parameters in the growth function $G(x', x)$ and the survival function $s(x)$. The major difference between the two growth functions is that the IPM model with the quadratic growth function predicted slower *Bd*-induced population declines than the linear model.

4.A.3 Loss of infection function: $l(x)$

The various models we fit for the loss of infection function $l(x)$ are given in Table 4.S2. Model 3 and Model 5 are the best models based on AIC. Model 5 in which temperature is a factor has a marginally lower AIC than Model 3 in which temperature is continuous. A likelihood ratio test shows that Model 5 does not provide an overwhelmingly better fit than Model 3 ($\chi^2_{df=1} = 3.676$, $p = 0.055$) so we used the Model 3 (the linear model) because it allowed us to interpolate over all temperatures between 4 and 20 °C.

When fitting Model 3, there were three highly influential data points in which individuals lost infections after having a log zoospore of 8.3, 10.36, and 8.1. Individuals with these large losses had similar pre-loss loads at the next swabbing event (Figure 4.S9), leading us to believe that these large losses were likely due to experimental error. Therefore, we excluded these points when fitting the model.

4.A.4 Initial infection burden function: $G_0(x')$

The various models we fit for the initial function burden function are given in Table 4.S3. The normalized residuals of the full model were not significantly different than a normal distribution (Shapiro-Wilk test for normality: $p = 0.827$), thereby justifying the assumption of normality for the initial infection distribution. Similar to the loss of infection function, there were three outlying log-zoospore

initial loads of 7.15, 8.26, and 11.81, which were the same spurious transitions observed in the loss of infection function, but in this case the points were an unrealistic gain in zoospores after the unrealistic loss of zoospores (Figure 4.S9). These points were again excluded from the analysis. After this exclusion, there was only one transition from 0 to infected at 20 °C. Diagnostic plots for the model used in the main text (given in bold in Table 4.S3) are given in Figure 4.S3.

4.A.5 Density-independent transmission function: $\phi(T)$

We explored three different density-independent transmission models. In the first model, the probability of infection was independent of temperature (Model 1). In the second model, temperature was a linear predictor of the probability of infection (Model 2). In the third model, temperature was a factor predicting the probability of infection. The model with a linear effect of temperature was the best model based on AIC criteria (Table 4.S4).

4.B The effect of eviction on the *Bd-Rana muscosa* Integral Projection Model

Given the parameterized density-independent IPM described in the main text, we examined the effects of eviction (loss of individuals from the model because their predicted future loads are outside the model range) using the examples and code given in Williams et al. (2012). In Table 4.S5, we show the maximum size-dependent eviction value $\epsilon(x)$ as given by equation 2 in (Williams et al., 2012) for the host-parasite IPM model at four different temperatures. These values are non-zero, indicating that eviction is occurring in our parameterized IPM with a lower bound of -5 and an upper bound of 18. To assess the effect of eviction on the IPM predictions, we also show the value $d\lambda$ which gives the effect of eviction on the predicted population growth rate (Williams et al., 2012). For all temperatures between 4 and 20 °C (4 temperatures shown in Table 4.S5), $d\lambda$ is very small indicating that despite eviction occurring in the parameterized IPM, it is having very little effect on the predictions of the IPM. Therefore, we felt confident in interpreting the IPM with the given upper and lower bounds.

4.C R_0 for host-parasite Integral Projection Models

4.C.1 Derivation of R_0 for IPMs

Calculating R_0 for Integral Projection Models (IPM)s is challenging because IPMs can be used to represent the dynamics of both microparasites and macroparasites. Therefore, R_0 will need to be computed and understood differently depending on the which type of parasite is being considered and the structure of the IPM. For microparasites, R_0 is defined as the average number of secondary infections produced by a typical infectious individual over its infective lifetime (Diekmann et al., 1990). For macroparasites, R_0 is defined as "the number of new female parasites produced by an average female parasite when there are no density-dependent constraints acting anywhere in the life cycle of the parasite" (Tompkins et al., 2002). We adopt a microparasite definition of R_0 for the remainder of this discussion, bearing in mind that a host-parasite IPM could easily be formulated such that the macroparasite definition of R_0 is more appropriate.

To define a microparasite R_0 for the host-parasite IPM described in the main text (equations 1 and 2), we start by considering density-dependent transmission such that the probability of becoming infected in a time step t is

$$\phi(I(x,t)) = 1 - \exp\left(\beta \int_L^U I(x,t)dx\right) \tag{4.16}$$

We then note that the host-parasite IPM model can be analogously stated as a (S)usceptible-(I)nfected-(S)usceptible model with a continuous $I(x)$ class. When analyzing the IPM model, it is standard practice to discretize the IPM into some number of n bins such that the IPM can be represented as a matrix model with a large number of classes (Coulson, 2012). This could be thought of as re-expressing the SIS model with a continuous I class as an S-I_1-I_2-I_3-...-I_n-S model with many discrete I classes. Using this discretized approach, R_0 can be calculated using the methods described in Allen and van den Driessche (2008) and Klepac and Caswell (2011). Following the notation of Klepac and Caswell (2011), the partial

matrix representation of the IPM that we use to calculate R_0 is given by

$$\begin{bmatrix} S \\ \\ \mathbf{I} \end{bmatrix}(t+1) = \begin{bmatrix} 0 & 0 \\ \\ M(\mathbf{I}) & N(\mathbf{I}) \end{bmatrix} \begin{bmatrix} S \\ \\ \mathbf{I} \end{bmatrix}(t) = m(\mathbf{I}(\mathbf{t}))S(t) + \mathbf{U}\mathbf{I}(\mathbf{t}) = \mathbf{I}(t+1) \tag{4.17}$$

where the top two entries are 0 because they are not needed when calculating R_0 (i.e. R_0 only depends upon individuals entering the infected classes or individuals that are already in the infected classes), not because they are actually 0 in the IPM model (Oli et al., 2006; Klepac and Caswell, 2011). $\mathbf{I}$ is a vector of length n that gives the various infected parasite load classes. $m(\mathbf{I})$ is a vector of length n where each element gives the probability of transitioning from class S (uninfected/susceptible) to an infected class with parasite load x_i where i is between 1 and n. We use this notation loosely as it is really the probability of transitioning to an infected class with a load in the interval $x_i \pm \Delta/2$ where x_i is the midpoint of this interval. Δ arises from using the midpoint rule to evaluate the IPM (Easterling et al., 2000). Each ith element of the vector $m(\mathbf{I})$ is given by

$$m_i(\mathbf{I}(t)) = s_0\phi(\mathbf{I}(t))G_0(x_i)\Delta \tag{4.18}$$

where s_0 represents the probability of an uninfected individual surviving and $G_0(x_i)$ is the probability density function of transitioning from uninfected (S) to infected with a load of x_i as defined in the main text. Δ is needed to convert the probability density $G_0(x_i)$ to a probability.

$\mathbf{U}$ is a n x n matrix that specifies the transition probabilities of infected individuals among different load classes. The element in the ith row and the jth column of the matrix is given by

$$u_{ij} = s(x_j)(1 - l(x_j))G(x_i, x_j)\Delta \tag{4.19}$$

which gives the probability of an individual in the jth load class surviving ($s(x_j)$), not losing its infection ($1 - l(x_j)$), and transitioning to the load class of x_i in a time step ($G(x_i, x_j)$).

To calculate R_0, we then linearize $\mathbf{I}(t+1)$ about a vector $\mathbf{n}^*$ which we set to be a host population with only susceptibles (Rohani et al., 2009; Klepac and Caswell, 2011), $N^* = [S^* \ \mathbf{0}]$ where $\mathbf{0}$ is a vector

of zeros of length n. We then compute the Jacobian matrix evaluated at $\mathbf{n}^*$

$$\mathbf{J} = \left.\frac{d\mathbf{I}(t+1)}{d\mathbf{I}(t)}\right|_{\mathbf{n}^*} \tag{4.20}$$

which allows us to compute R_0 (Klepac and Caswell, 2011).

In the above case, one could compute $\mathbf{J}$ as follows. First compute, $\left.\frac{d\mathbf{UI}(t)}{d\mathbf{I}(t)}\right|_{\mathbf{n}^*}$ which is simply $\mathbf{U}$. This is just the transition matrix for the infected individuals of various load classes. Next, compute $\left.\frac{dm(\mathbf{I}(t))}{d\mathbf{I}(t)}\right|_{\mathbf{n}^*}$, which results in a column vector $\mathbf{m}$ of length n where each element is given by

$$\frac{dm_i(\mathbf{I}(t))}{d\mathbf{I}(t)} = \beta s_0 S^* G_0(x_i)\Delta \tag{4.21}$$

Now let $\mathbf{M}$ be an n by n matrix with each column being equal to $\mathbf{m}$ so that $\mathbf{J} = \mathbf{M} + \mathbf{U}$. R_0 is then given by

$$R_0 = \max \text{ eig}(\mathbf{M}(\mathbb{1} - \mathbf{U})^{-1}) \tag{4.22}$$

where $\mathbb{1}$ is the identity matrix and $\mathbf{M}$ is equivalent to the "fertility" matrix described in Klepac and Caswell (2011). "max eig" refers to the maximum eigenvalue of this matrix.

A helpful approximation of this result can be derived by "collapsing" the various infected classes $\mathbf{I}(t)$ into a single infected class $I(t)$. The model is then reduced to a simple SIS model with the following transition matrix (where we again include 0s where the transitions do not affect the calculation of R_0)

$$\begin{bmatrix} S \\ \\ I \end{bmatrix}(t+1) = \begin{bmatrix} 0 & 0 \\ \\ s_0\phi(I(t)) & \bar{s}_I(1-\bar{l}) \end{bmatrix}\begin{bmatrix} S \\ \\ I \end{bmatrix}(t) = s_0\phi(I(t))S(t) + \bar{s}_I(1-\bar{l})I(t) = I(t+1) \tag{4.23}$$

where $\bar{s}_I$ is the survival probability for an average infected individual and $\bar{l}$ is the probability of an average infected individual losing an infection.

Using the sames steps as above the resulting value of R_0 is

$$R_0 = \frac{\beta s_0 S^*}{1 - \bar{s}_I(1-\bar{l})} \tag{4.24}$$

4.C.2 Application of R_0 to *Bd-R. muscosa*

Using equations 4.22 and 4.24, we computed R_0 for the *Bd-Rana muscosa* system described in the main text as an illustrative example. Note that in this example, we assumed density dependent transmission without any probability of acquiring an infection from the environmental reservoir as we do in the main text. We make this simplification here because without accounting for the decay of the pathogen in the environment, an R_0 that accounts for both transmission due to the environment and other infected individuals would be trivially ∞ (Rohani et al., 2009). An environmental reservoir could be more explicitly incorporated in the host-parasite IPM by including an additional state variable $Z(t)$ which gives the total number of parasites in the environment at time t.

We set the transmission coefficient $\beta = 9.82e10^{-4}$ which was the transmission coefficient estimated in Rachowicz and Briggs (2007) for density-dependent transmission in *Bd-Rana muscosa* and assumed an initial susceptible population of 100 frogs ($S^* = 100$). Otherwise, all values for the hosts-parasite IPM were as given in Table 1 the main text. To compute $\bar{s}_I$ and $\bar{l}$ in equation 4.24, we assumed a density-independent host-parasite IPM (equations 7 and 8 in the main text) with $\phi = 1-\exp(-\beta)$ and calculated the stable parasite load distribution conditional on infection ($p(x)$) giving the probability density of having some parasite load x. We used this probability distribution to compute the expected survival and loss probability of an average infected individual as $\bar{s}_I = \int_L^U s(x)p(x)dx$ and $\bar{l} = \int_L^U l(x)p(x)dx$, respectively.

Figure 4.S10 shows the temperature dependence of R_0 for this illustrative example parameterized from the *Bd-R. muscosa* IPM. Notice that the approximation given by equation 4.24 is nearly identical to the predictions for R_0 from equation 4.22. At low temperatures, R_0 is less than 1 and proceeds to increase as temperature increases. At approximately 12 °C, R_0 ¿ 1. However, around 17 °C the R_0 reaches a maximum and begins to decline. This is due to the average probability of losing an infection $\bar{l}$ quickly and non-linearly decreasing as temperature increases and the average probability of surviving with an infection $\bar{s}_I$ holding relatively constant with temperature and then rapidly decreasing as temperature increases past 17 °C.

Table 4.S1: Candidate models for the growth function $G(x', x)$. All models assumed a normal distribution for the response variable. T is temperature, x is log zoospore size at time t, and r_i represents a random effect of an individual frog. The model with the bold AIC value is the model used in the main text.

Model	**Mean Component**	**Variance Component**	**AIC**
1	$\mu(x,T) = b_0 + b_1 x + b_2 T$	σ^2	1067.8
2	$\mu(x,T) = b_0 + b_1 x + b_2 T$	$\sigma^2(x) = \nu \exp(2cx)$	**1060.0**
3	$\mu(x,T) = b_0 + b_1 x + b_2 T$	$\sigma^2(x,T) = \nu \exp(2c_1 x + 2c_2 T)$	1060.7
4	$\mu(x,T)_i = b_0 + b_1 x + b_2 T + r_i$	$\sigma^2(x) = \nu \exp(2cx)$	1061.8
5	$\mu(x,T) = b_0 + b_1 x + b_2 T + b_3 xT$	$\sigma^2(x) = \nu \exp(2cx)$	1061.0
6	$\mu(x,T) = b_0 + b_1 x + b_2 T + b_3 xT + b_4 x^2$	$\sigma^2(x) = \nu \exp(2cx)$	1045.3
7	$\mu(x,T) = b_0 + b_1 x + b_2 T + b_3 x^2$	$\sigma^2(x) = \nu \exp(2cx)$	1049.3

Table 4.S2: Candidate models for the loss of infection function $l(x)$. All models assumed a binomial distribution for the response variable. T is temperature, T_i is temperature as a factor, and x is log zoospore load at time t. The model with the bold AIC value is the model used in the main text.

Model	**Mean Component**	**AIC**
1	$\text{logit}(l(x)) = b_0 + b_1 x$	193.0
2	$\text{logit}(l(T)) = b_0 + b_1 T$	202.1
3	$\text{logit}(l(x,T)) = b_0 + b_1 x + b_2 T$	**180.4**
4	$\text{logit}(l(x,T)) = b_0 + b_1 x + b_2 T + b_3 x T$	182.4
5	$\text{logit}(l(x,T))_i = b_0 + b_1 x + T_i$	178.7

Table 4.S3: Candidate models for the initial infection burden function $G_0(x')$. All models assumed a normal distribution for the response variable. T is temperature, T_i is temperature as a factor, and x is log zoospore load at time t. The model with the bold AIC value is the model used in the main text.

Model	**Mean Component**	**Variance Component**	**AIC**
1	$\mu(T) = b_0 + b_1 T$	σ^2	149.2
2	$\mu(T)_i = b_0 + T_i$	σ^2	148.5
3	$\mu(T) = b_0 + b_1 T$	$\sigma^2(T) = \nu \exp(2cT)$	**145.2**

Table 4.S4: Candidate models for the density-independent transmission function $\phi(T)$. All models assumed a binomial distribution for the response variable. T is temperature and T_i is temperature as a factor. The model with the bold AIC value is the model used in the main text.

Model	**Mean Component**	**AIC**
1	$\text{logit}(\phi) = b_0$	198.0
2	$\text{logit}(\phi(T)) = b_0 + b_1 T$	**193.9**
3	$\text{logit}(\phi(T)_i) = b_0 + T_i$	195.8

Table 4.S5: Table shows the effect of eviction on the *Batrachocytrium dendrobatidis-Rana muscosa* Integral Projection Model described in the main text at 4 different temperatures. $\epsilon(x)$ specifies the maximum value of eviction occurring in the IPM model as given by equation 2 in Williams et al. (2012). A value of zero indicates no eviction is occurring while a non-zero value indicates that eviction is occurring in the IPM. $d\lambda$ gives the effect of this eviction on the predicted population growth rate. In other words, how much would this growth rate change if no eviction was occurring. Despite eviction occurring in the *Bd-R. muscosa* IPM, it is having little effect on the predicted population growth rate.

***Bd-R. muscosa* IPM**	$\epsilon(x)$	$d\lambda$
at 4 °C	0.32	2.26e-05
at 10 °C	0.26	2.74e-05
at 15 °C	0.22	2.52e-05
at 20 °C	0.18	1.45e-08

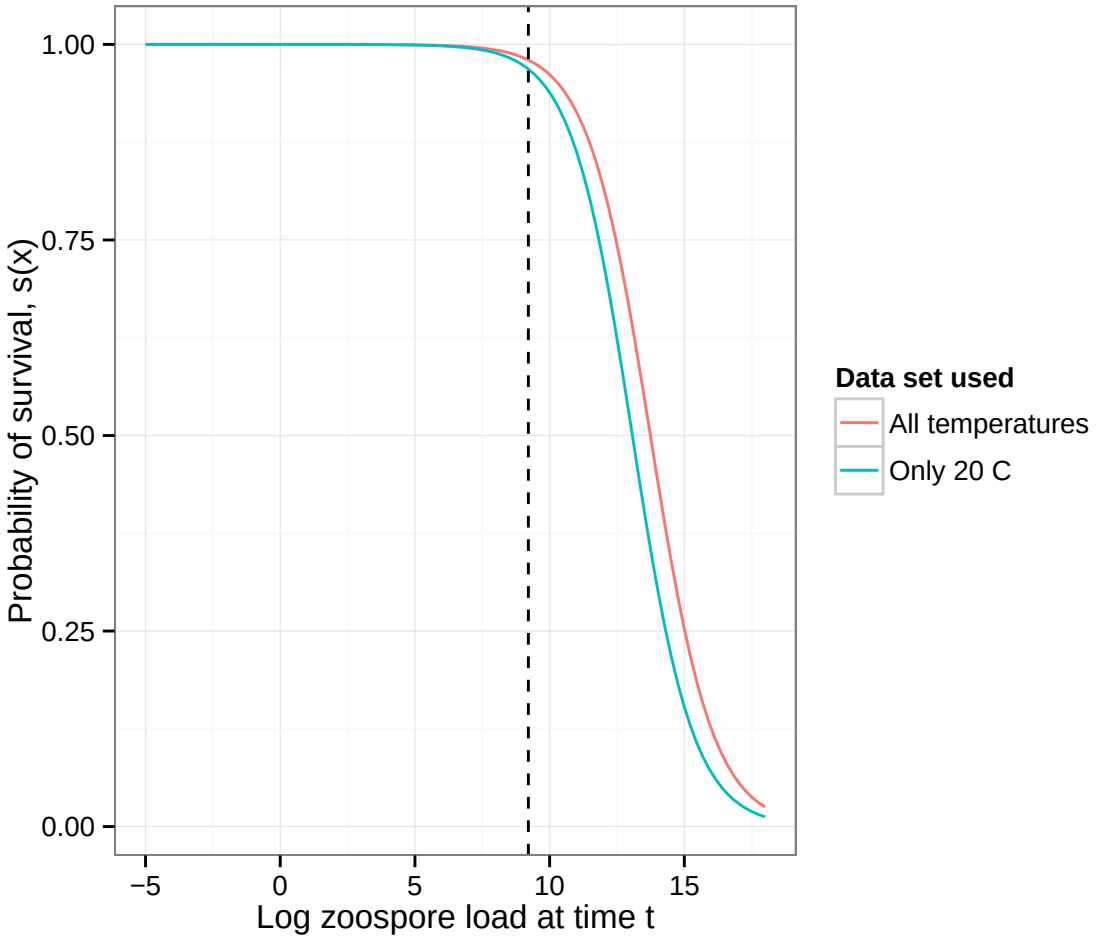


Figure 4.S1: Comparison of survival functions fit from two different subsets of the data. The blue line shows the survival function used in the Integral Projection Model (IPM) analysis described in the main text and only includes data from individuals housed at 20 °C. The red line shows an alternative survival function that was parameterized using the data from all temperatures used in the experiment (4, 12, 20 °C). The dashed vertical line gives the 10,000 zoospore threshold reported by Vredenburg et al. (2010), which gives an approximate threshold at which *R. muscosa* begins to experience *Bd*-induced mortality in the field. While the survival curves from the two models are very similar, our elasticity analysis shows that even this small difference can have large effects on whether an *R. muscosa* population can persist through an epizootic at high temperatures.

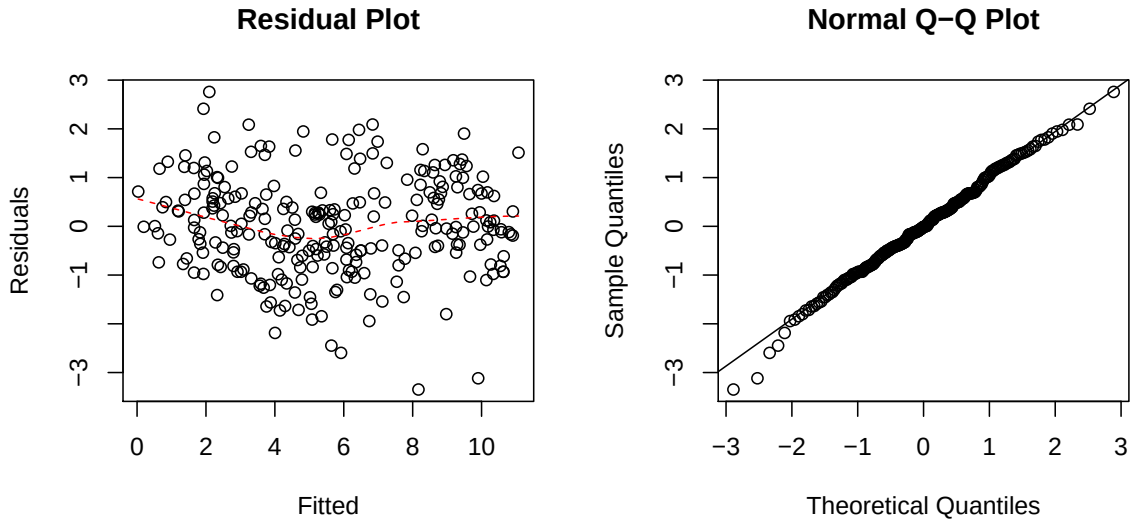


Figure 4.S2: Diagnostic plots for the Model 2 in Table 4.S1. The noticeable pattern in the residual plot (red line) can be accounted for with a quadratic term in the growth function (Model 6, Table 4.S1). As discussed in the subsection *Growth Function:* $G(x', x)$ we chose to use this linear model for the growth function, but explored the effects of the alternative, non-linear growth function on the IPM predictions.

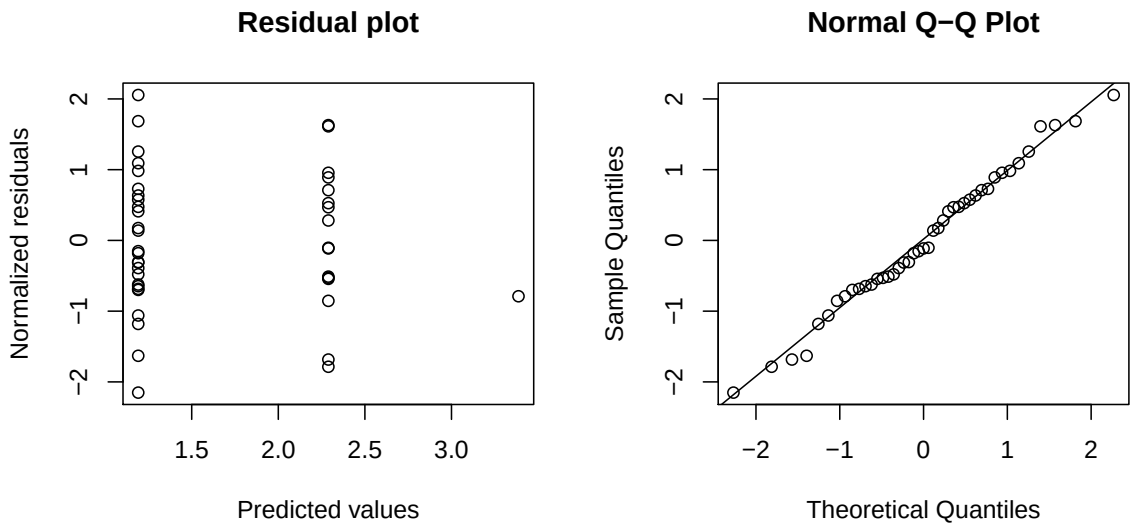


Figure 4.S3: Diagnostic plots for the Model 3 in Table 4.S3. The data point to the far right in the residual plot shows the single data point for a transition of an individual from 0 to infected at a temperature of 20 °C.

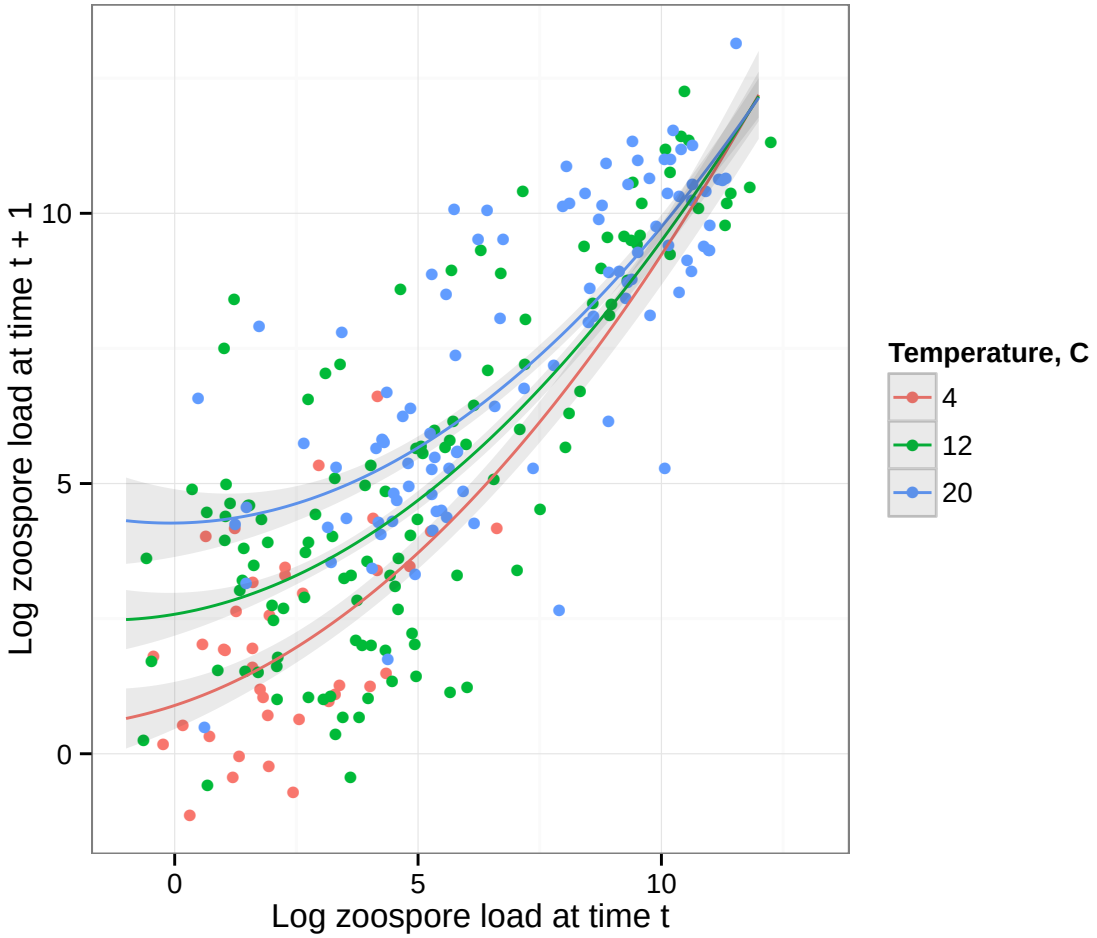


Figure 4.S4: The best fit quadratic empirical growth function given by Model 6 in Table 4.S1. As log-zoospore load at time t decreases the growth function flattens and eventually begins to increase for small enough values of log-zoospore load at time t.

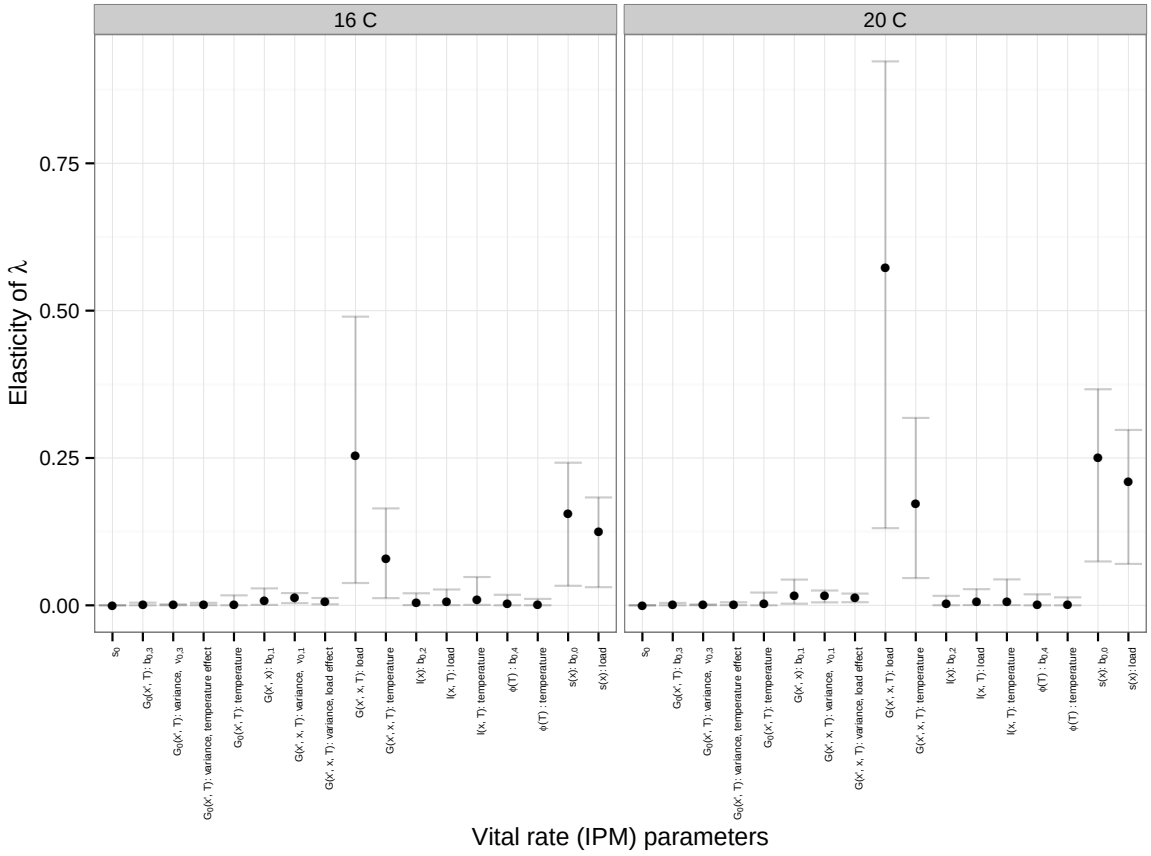


Figure 4.S5: A local elasticity analysis of the population growth rate λ to the vital rate parameters used in the *Bd-Rana muscosa* IPM. The x axis gives all the vital rate parameters used in the *Bd-R. muscosa* IPM model. Each x axis label specifies the vital rate function to which a parameter belongs as well as the identity of that parameter. The parameters labeled as $b_{0,j}$ represent the intercepts of the given vital rate functions. The parameters labeled as load and temperature identify the load and temperature parameters of the given vital rate function. The parameters specified as variance refer to the parameters affecting the variance of the vital rate function, where $\nu_{0,j}$ gives the variance of the vital rate function when the effect of other covariates on the variance is 0. The points represent the median elasticity of λ to a given parameter based on 1000 simulations and the error bars give the first and third quartiles of the uncertainty around this elasticity.

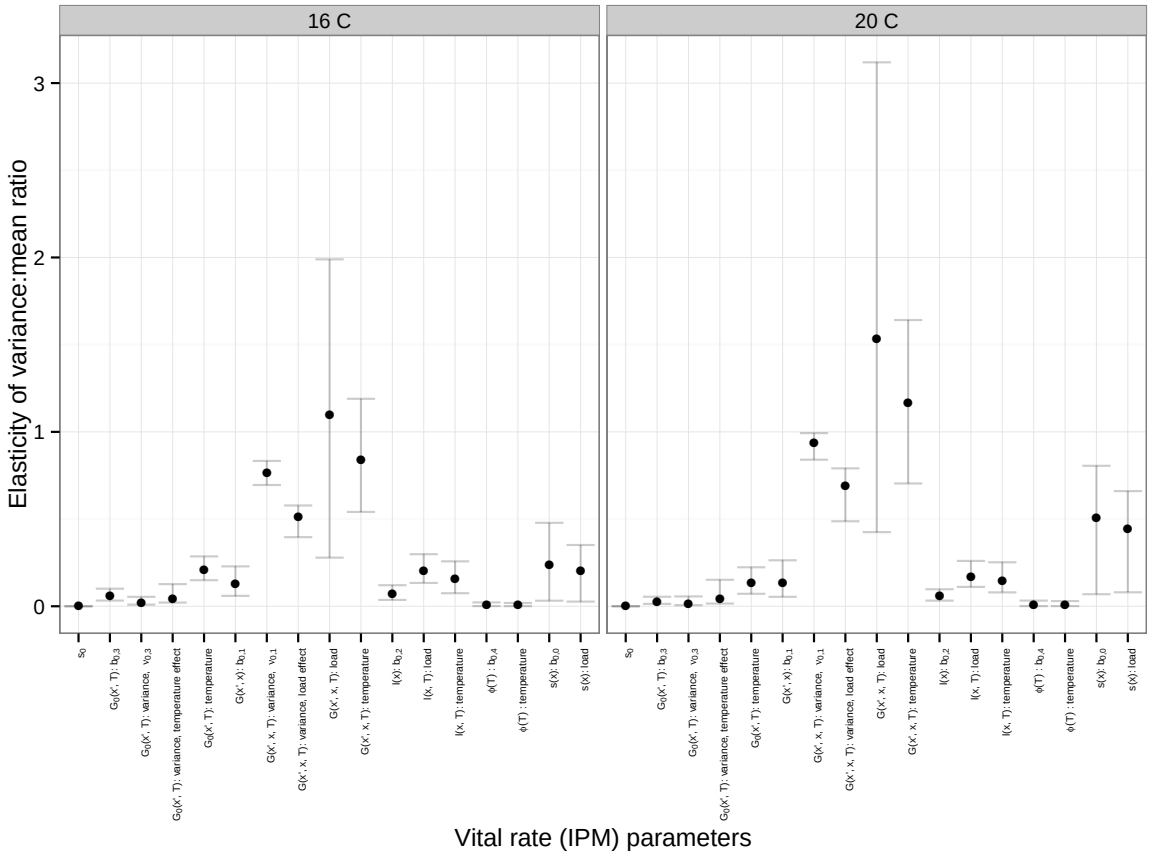


Figure 4.S6: A local elasticity analysis of the variance to mean ratio of the *Bd* load distribution to the vital rate parameters used in the *Bd-Rana muscosa* IPM. The x axis gives all the vital rate parameters used in the *Bd-R. muscosa* IPM model. Each x axis label specifies the vital rate function to which a parameter belongs as well as the identity of that parameter. The parameters labeled as $b_{0,j}$ represent the intercepts of the given vital rate functions. The parameters labeled as load and temperature identify the load and temperature parameters of the given vital rate function. The parameters specified as variance refer to the parameters affecting the variance of the vital rate function, where $\nu_{0,j}$ gives the variance of the vital rate function when the effect of other covariates on the variance is 0. The points represent the median elasticity of the variance to mean ratio to a given parameter based on 1000 simulations and the error bars give the first and third quartiles of the uncertainty around this elasticity.

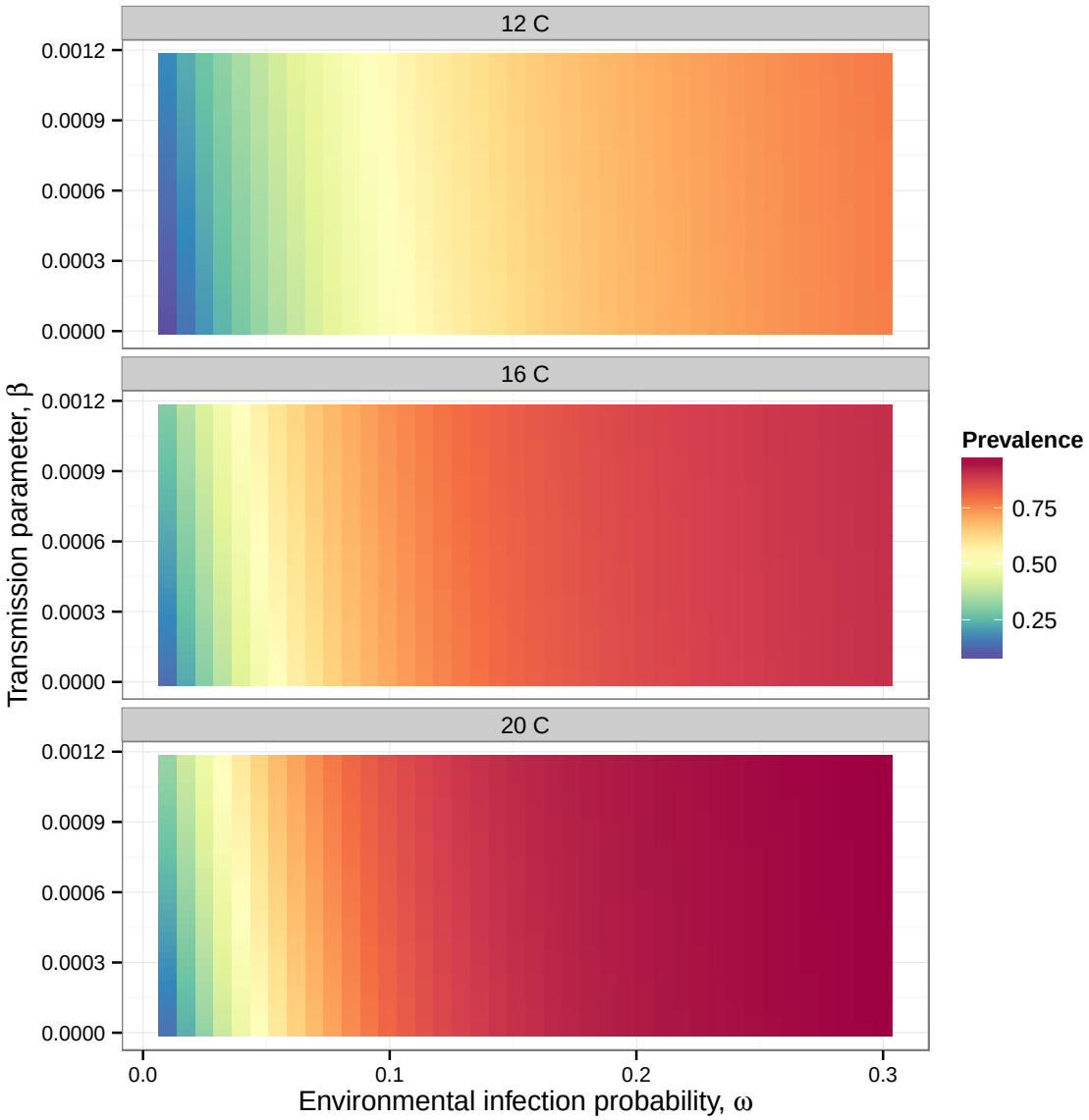


Figure 4.S7: Prevalence of *Bd* at the end of a 120 day epizootic given different transmission coefficients (β) and environmental infection probabilities (ω) for the density-dependent transmission function. This plot shows 40 x 40 systematically chosen pairs of β and ω for which the *Bd*-prevalence dynamics were examined. Each panel shows the change in *Bd* prevalence in *Rana muscosa* populations with the different parameter combinations for a given temperature between 12 and 20 °C.

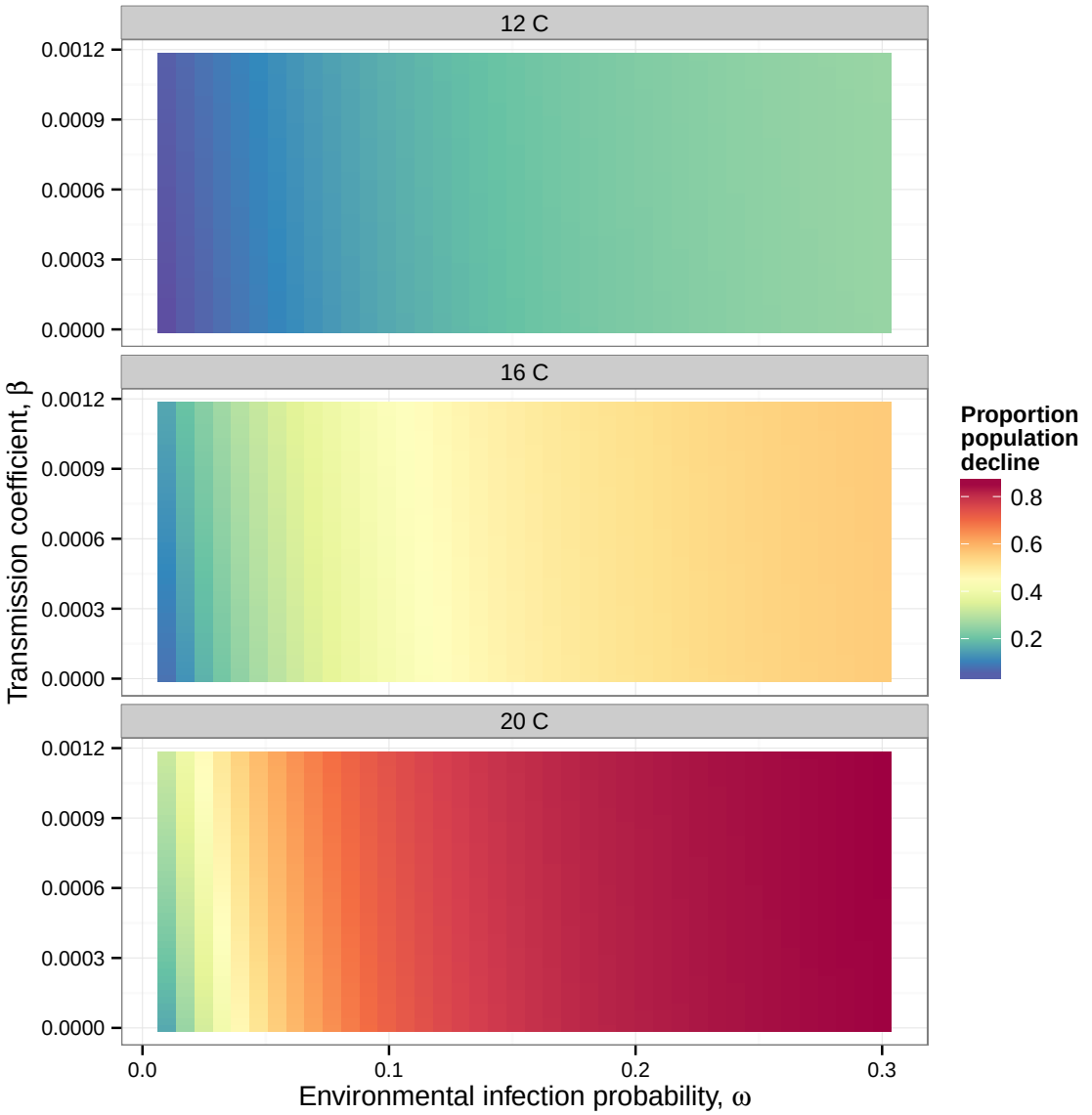


Figure 4.S8: Proportional population loss of *Rana muscosa* at the end of a 120 day epizootic given different transmission coefficients (β) and environmental infection probabilities (ω) for the density-dependent transmission function. This plot shows 40 x 40 systematically chosen pairs of β and ω for which the population dynamics were examined. Each panel shows the change in proportional population loss for *R. muscosa* with the different parameter combinations for a given temperature between 12 and 20 °C.

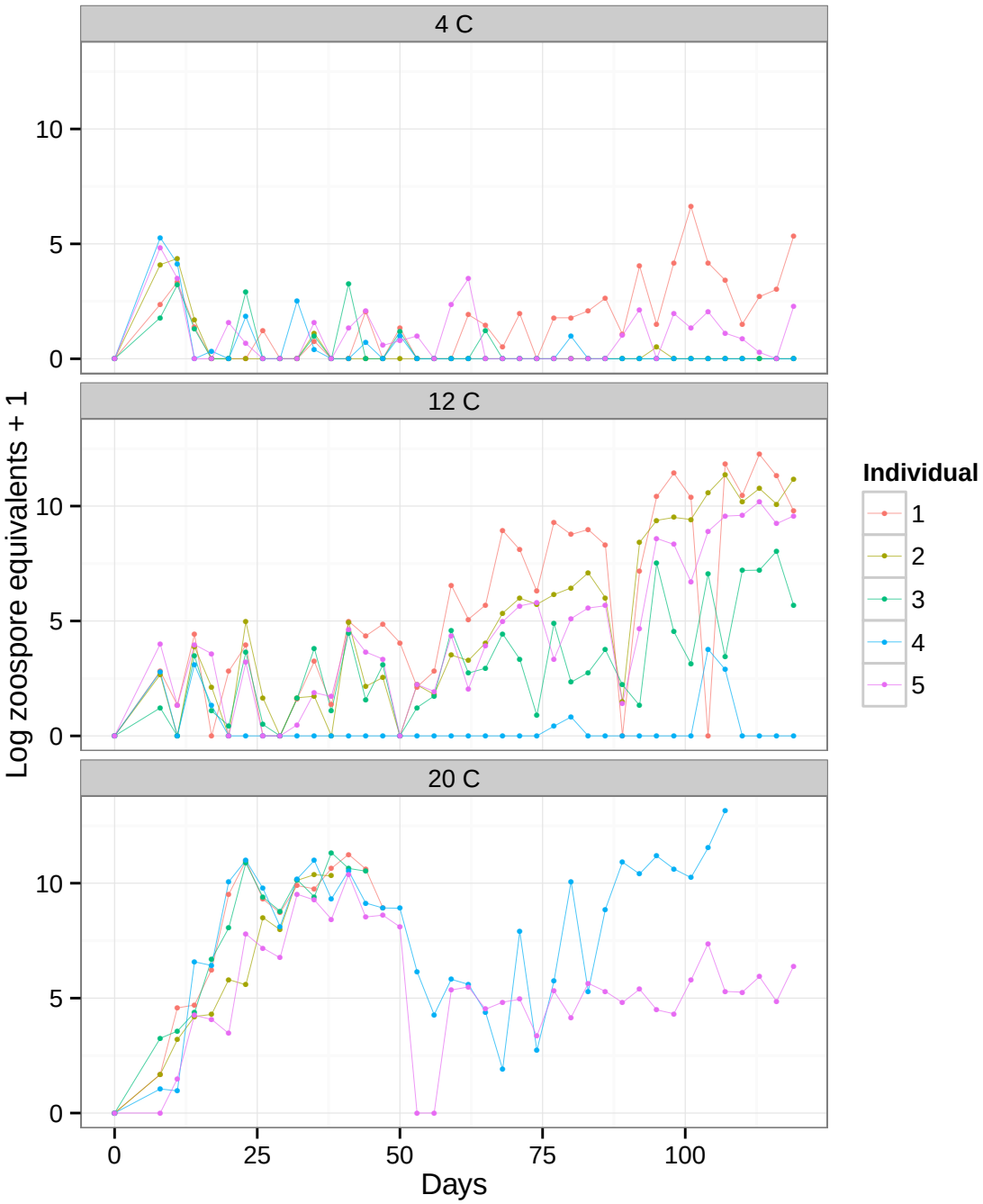


Figure 4.S9: Infection trajectories of individual *Rana muscosa* housed at 4, 12, and 20 °C. Each line represents the *Bd* load trajectory of a particular *Rana muscosa* individual.

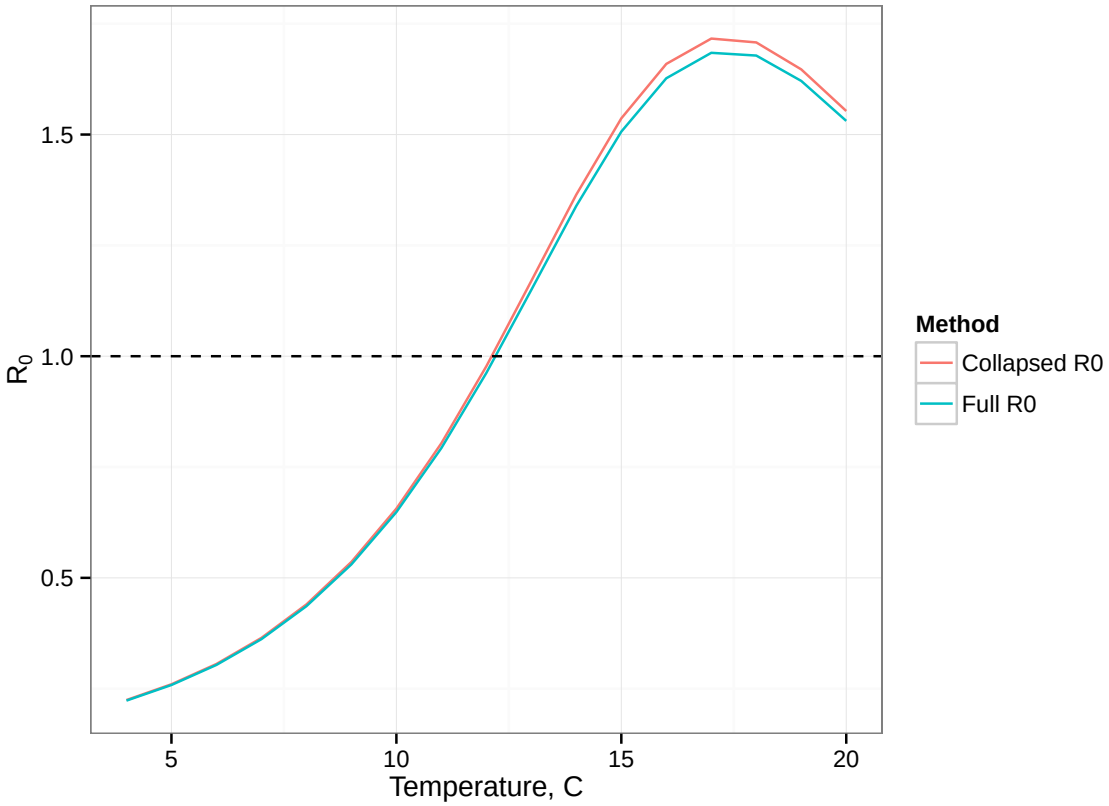


Figure 4.S10: Plots of R_0 for a *Rana muscosa-Bd* IPM with density dependent transmission parameterized based on the parameters provided in Table 1 in the main text. The only parameter that was not from Table 1 was the transmission coefficient β which was set to be $9.82e10^{-4}$ based on Rachowicz and Briggs (2007). The number of susceptible individuals in the initial population was set to $S^* = 100$. R_0 was computed using both equation 4.22 (the blue line, Full R_0) and equation 4.24 (the red line, Collapsed R_0)

Chapter 5

Resistance, tolerance and environmental transmission dynamics determine host extinction risk in a load-dependent amphibian disease

5.1 Abstract

While disease-induced extinction is generally considered rare, a number of recently emerging infectious diseases with load-dependent pathology have led to extinction in wildlife populations. Transmission is a critical factor affecting disease-induced extinction, but the relative importance of transmission compared to load-dependent host resistance and tolerance is currently unknown. Using a combination of models and experiments on an amphibian species suffering extirpations from the fungal pathogen *Batrachochytrium dendrobatidis* (Bd), we show that while transmission from an environmental Bd reservoir increased the ability of Bd to invade an amphibian population and the extinction risk of that population, Bd-induced extinction dynamics were far more sensitive to host resistance and tolerance than to Bd transmission. We demonstrate that this is a general result for load-dependent pathogens, where non-linear resistance and tolerance functions can interact such that small changes in these functions lead to drastic changes in extinction dynamics.

5.2 Introduction

Disease-induced extinction of a host population is considered rare (De Castro and Bolker, 2005; Smith et al., 2006; McCallum, 2012). This is because in many systems disease transmission is an increasing function of the density of infected hosts (i.e. density-dependent transmission) such that decreasing host density reduces disease transmission and prevents disease-induced extinctions (McCallum and Dobson, 1995; Gerber et al., 2005). Theoretical models suggest that to drive a host population extinct, a disease needs to have alternative transmission dynamics such that declining host density does not prevent further disease transmission (De Castro and Bolker, 2005; Smith et al., 2006; McCallum et al., 2009). For example, frequency-dependent transmission, in which hosts have a density-independent number of contacts with other hosts per unit time (McCallum et al., 2001), or abiotic/biotic reservoirs for the pathogen are two transmission scenarios that can lead to disease-induced host extinction. In both cases, decreasing host density does not necessarily lead to a decrease in disease transmission. Despite the rarity of disease-induced host extinction (Smith et al., 2006), a number of wildlife diseases, such as chytridiomycosis in amphibians, white-nose syndrome in bats, and facial tumor disease in Tasmanian devils, have recently been identified as the causative agents of host declines and population extinctions

(Skerratt et al., 2007; Blehert et al., 2008; McCallum et al., 2009).

While the characteristics of the transmission function in these emerging infectious diseases may ultimately determine whether a population experiences disease-induced extinction (McCallum, 2012), extinction dynamics will also be influenced by how resistant (i.e. the ability of a host to reduce or eliminate a pathogen conditional on exposure; Boots et al., 2009; Medzhitov et al., 2012) and/or tolerant (i.e. the ability of a host to persist with a pathogen load that is typically lethal for non-tolerant individuals; Roy and Kirchner, 2000; Medzhitov et al., 2012) a host is to the pathogen. Therefore, when managing disease-induced declines and extinctions, it may be important to manage not only for the transmission dynamics, but also the level of host tolerance and resistance in a population (Kilpatrick, 2006; Langwig et al., 2015, 2017; Epstein et al., 2016). However, the conditions under which it might be more effective to manage for resistance and tolerance instead of transmission are currently unknown.

Understanding the relative importance of resistance and tolerance compared to transmission in driving extinction dynamics has important implications for managing the emerging amphibian disease chytridiomycosis. Chytridiomycosis is caused by the amphibian chytrid fungus *Batrachochytrium dendrobatidis* (Bd) and has resulted in widespread amphibian declines and extinctions (Daszak et al., 2003; Skerratt et al., 2007). Bd is a cutaneous fungus that disrupts the osmoregulatory ability of amphibian skin, leading to the potentially fatal disease chytridiomycosis (Voyles et al., 2007, 2009). While a number of transmission-related factors, including an environmental pool of Bd zoospores and biotic reservoirs, are hypothesized to contribute to disease-induced extinction of amphibian populations (Rachowicz and Briggs, 2007; Mitchell et al., 2008; Briggs et al., 2010; McCallum, 2012; Doddington et al., 2013), few studies have attempted to quantify the transmission function itself (but see Rachowicz and Briggs (2007) and Bielby et al. (2015)). Quantifying the transmission function is important when considering disease-induced extinction because it determines the ability of a pathogen to invade a population as well as drive a host population extinct (De Castro and Bolker, 2005; Gerber et al., 2005).

In addition to the transmission function, extinction dynamics in amphibian-Bd systems also depend on the dynamics of Bd load on an amphibian host (Briggs et al., 2010). Varying Bd load dynamics among amphibian populations, potentially induced by varying resistance and tolerance mechanisms, can promote population-level persistence following epizootics (Retallick et al., 2004; Briggs et al., 2010; Grogan et al., 2016; Savage and Zamudio, 2016). This variation in resistance and tolerance could be due to a number of different mechanisms including innate and acquired immune responses (Ellison et al.,

2015), differences in host susceptibility (Knapp et al., 2016), variation in the amphibian microbiome (Harris et al., 2009; Jani and Briggs, 2014), temperature-dependent Bd growth (Forrest and Schlaepfer, 2011; Knapp et al., 2011), and variable virulence in Bd strains (Rosenblum et al., 2013; Jenkinson et al., 2016). Load dynamics are important in Bd systems because disease-induced mortality of amphibians is highly load-dependent, with survival probability sometimes decreasing rapidly at high Bd loads (Stockwell et al., 2010; Vredenburg et al., 2010). This attribute of load-dependent survival leads to a simple, but largely untested hypothesis in host-pathogen systems: host populations that are either able to prevent large increases in pathogen load (via resistance mechanisms) or tolerate high pathogen loads (via tolerance mechanisms), will experience reduced disease-induced extinction risk, even when the transmission rate is high. This is a general hypothesis for load-dependent wildlife diseases and amphibian-Bd interactions provide an ideal system in which to test it.

The above hypothesis can be phrased as the following question: How important is transmission compared to host tolerance and resistance for mitigating disease-induced extinction? Answering this question requires quantifying the transmission function, something rarely done in amphibian-Bd systems (Kilpatrick et al., 2010). Moreover, understanding the role of this transmission function in the ability of Bd to invade an amphibian population will be important for accurately understanding any subsequent Bd-induced extinctions (Gerber et al., 2005). Therefore, we also ask two additional questions: What is the nature of the transmission function in amphibian-Bd systems? How does this transmission function affect the ability of Bd to invade an amphibian population? We use a combination of experiments and dynamical models to show that empirical patterns of Bd transmission are best modeled using an environmental Bd pool and that this pool significantly increases the ability of Bd to invade a population and the population-level extinction risk. However, despite this large effect of the environmental pool, we show that host resistance and tolerance are far more influential on Bd-induced extinction dynamics than transmission. This is likely a general property of load-dependent diseases in which non-linear resistance and/or tolerance functions interact such that managing for resistance and tolerance can more effectively mitigate disease-induced extinction than managing for transmission.

5.3 Methods

To answer the questions posed above, we focused on Bd-induced extinction dynamics in the Mountain yellow-legged frog complex (*Rana muscosa* and *Rana sierrae*, henceforth *R. muscosa*). *R. muscosa* are native to California's Sierra Nevada mountains and have experienced significant Bd-induced population declines and extinctions over the last four decades (Briggs et al., 2005; Vredenburg et al., 2010; Briggs et al., 2010; Knapp et al., 2016). Using this host-parasite system, the *Methods* section is organized as follows.

First, we used a laboratory experiment to quantify the well-known importance of temperature-dependent Bd growth dynamics on *R. muscosa* (Andre et al., 2008; Wilber et al., 2016). Second, we used a mesocosm experiment to quantify the nature of the transmission function in the *R. muscosa*-Bd system, testing for both density-dependent transmission, frequency-dependent transmission, and transmission from an environmental zoospore pool. Third, we used the results from these experiments to build a discrete-time, host-parasite Integral Projection Model (IPM) and derived R_0 with an environmental zoospore pool. We used this result to explore the effects of the zoospore pool on Bd invasion. Finally, to answer our primary question regarding the relative importance of transmission compared to resistance and tolerance on extinction risk, we extended our model to consist of a within-year component in which Bd transmission and disease-induced amphibian mortality occurred and a between-year component in which *R. muscosa* demographic transitions occurred (Figure 5.1). Using this hybrid model, we explored the sensitivity of Bd-induced extinction to transmission, resistance, and tolerance.

5.3.1 Laboratory and mesocosm experiments

We used laboratory and mesocosm experiments to quantify the temperature-dependent load dynamics and the transmission dynamics in the *R. muscosa*-Bd system. The laboratory experiment is fully described in Wilber et al. (2016) and consisted of 15 adult frogs housed at 3 different temperatures (4 °C, 12 °C, 20 °C; 5 frogs per temperature). Wilber et al. (2016) used this experiment to parameterize four functions relating to Bd load dynamics: a load-dependent host survival function and a temperature-dependent Bd growth function, loss of infection function, and initial infection function (Fig. 5.S1, see Table 1 for function descriptions).

To quantify the nature of the transmission function in this system, we performed a mesocosm

experiment that consisted of four different density treatments: 1, 4, 8, and 16 uninfected adult frogs per mesocosm (volume ≈ 1 m^3). Each treatment was replicated four times for a total of 16 mesocosms (see Section 5.A). In addition to the uninfected adults, each mesocosm started with five infected tadpoles, which could release Bd zoospores into the environment and subsequently infected adults. All of the adults in a mesocosm were uniquely identifiable by pit tags, but the five tadpoles were not. The experiment ran for 74 days and every 4-8 days all adults and tadpoles in a mesocosm were swabbed and the zoospore load on each was determined using quantitative PCR (Boyle et al., 2004). Frogs and tadpoles within a mesocosm were always swabbed on the same day.

To estimate the transmission function, we measured the load transitions on all adult frogs from time t to $t + \Delta t$ over the first 32 days of the experiment (Δt = 4-8 days depending on the time between swabs in the experiment). We only used the first 32 days of the experiment as after this time point all amphibians experienced an unexplained decline in zoospore loads (see Fig. 5.S2). However, because the load trajectories over these first 32 days were consistent with other experiments (e.g. Wilber et al., 2016) and transitions from uninfected to infected tended to occur before day 32, we felt confident in estimating transmission dynamics from only the first 32 days. Moreover, we replicated the experiment *in silico* to demonstrate that we could recover known transmission functions over the first 32 days of the experiment (Section 5.A). As transmission is the probability of an uninfected individual gaining an infection in a time step, we only included transitions where Bd load was 0 at time t ($n = 333$). If the load at time $t + \Delta t$ was positive we assigned this data point a value of 1 (infected) and if the load was still zero we assigned it a value of 0 (uninfected).

Uninfected frogs can acquire Bd infection through contact with other infected frogs and through contact with zoospores in an environmental Bd pool (Courtois et al., 2017). To account for these different pathways, we fit two sets of transmission models to the data. The first set of models assumed a constant level of infection from the zoospore pool and either density-dependent or frequency-dependent transmission from conspecifics (Table 2). The second set of models allowed transmission to be a function of how many zoospores were in the zoospore pool at time t in addition to either density-dependent or frequency-dependent transmission. In Section 5.A, we describe how we defined and fit our transmission models with a dynamic zoospore pool. In short, we used the transmission function $\phi(t) = 1 - \exp(-\Lambda(Z(t), I(t))\Delta t)$ and allowed for the zoospore pool $Z(t)$ at time t to be an unobserved, dynamic variable that lost zoospores due to environmental decay and gained zoospores

due to production from infected adults and tadpoles at every time step. $I(t)$ is the number of infected adults at time t. Table 2 gives the transmission functions and the resulting fits to the data from the mesocosm experiment.

5.3.2 The host-parasite IPM and R_0

The host-parasite IPM

Using the aforementioned laboratory and mesocosm experiments, we parameterized a host-parasite Integral Projection Model (IPM) where Bd load on an individual frog was the continuous trait being modeled (Fig. 5.1B; Metcalf et al., 2015; Wilber et al., 2016). Bd load on a frog is estimated as the number of copies of Bd DNA detected on standardized skin swabs via quantitative PCR (Boyle et al., 2004) and provides a continuous measure of infection intensity between 0 (uninfected) and an arbitrarily large Bd infection. The IPM is a discrete time model and here a single time step is three days. This time step is on the same scale as the generation of time of Bd, which ranges between 4-10 days depending on temperature (Piotrowski et al., 2004; Woodhams et al., 2008). We used the discrete-time IPM because it is easily parameterized from laboratory data which is collected at discrete time intervals.

This IPM tracks two discrete stages at time t: the density of susceptible adults $S_A(t)$ in the population and the density of zoospores in the environment $Z(t)$. This model also tracks a continuous, infected stage $I_A(x,t)$ where x is ln Bd load and $\int_{L_x}^{U_x} I_A(x,t)dx$ gives the density of adult frogs with a ln Bd load between a lower bound L_x and an upper bound U_x at time t. This continuous, infected stage tracks the distribution of Bd loads at any time t in the population.

Considering these discrete and continuous stages, the amphibian-Bd IPM can be written as follows (Fig. 5.1B)

$$S_A(t+1) = S_A(t)s_0(1-\phi) + \int_{L_x}^{U_x} I_A(x,t)s_A(x)l_A(x)dx \tag{5.1}$$

$$I_A(x',t+1) = \int_{L_x}^{U_x} I_A(x,t)s_A(x)(1-l(x))G(x',x)dx + S_A(t)s_0\phi G_0(x') \tag{5.2}$$

$$Z(t+1) = Z(t)\nu + \mu_A \int_{L_x}^{U_x} \exp(x)I_A(x,t)dx - \psi(S_A(t), Z(t)) \tag{5.3}$$

$S_A(t+1)$ describes the density of susceptible adults at time $t+1$ and is determined by the number

of adults that survive and do not become infected in a time step (first term in equation 1) and the number of infected adults that survive and lose their infection in a time step (second term in equation 1). $I_A(x', t+1)$ describes the density of infected adults with a ln Bd load x' at time $t+1$ and is determined by infected adults who survive with load x, do not lose their load x, and experience a change in load from x to x' in a time step (first term in equation 2) and from uninfected adults who survive, become infected, and gain an initial Bd load of x' in a time step (second term in equation 2). The vital rate functions contained in equations 1-3 are described in Fig. 5.1B and Table 1.

The equation $Z(t+1)$ gives the density of zoospores in the zoospore pool at time $t+1$. $Z(t+1)$ depends on three distinct terms: the survival probability of the zoospores in the environment from time t to $t+1$ (ν), contribution of zoospores from infected adults where μ_A is the proportion of total zoospores on adults contributed to the zoospore pool over a time step, and removal of zoospores from the zoospore pool by frogs transitioning from uninfected to infected. This removal term $\psi(S_A(t), Z(t))$ had very little effect on the dynamics of the system and we do not consider it further.

Based on the laboratory experiment described above and known Bd life history (Woodhams et al., 2008), we allowed various vital rate functions to be temperature-dependent (e.g. the survival of free-living zoospores, Table 1, Fig. 5.S1).

R_0 with an environmental reservoir

Using the IPM described in equations 1-3, we calculated R_0 to quantify how temperature and the transmission dynamics affected the ability of Bd to invade a *R. muscosa* population – a necessary condition for Bd-induced population declines and extinction. R_0 describes the average number of secondary infections produced over the lifetime of an average infected agent (Diekmann et al., 1990; Dietz, 1993). When $R_0 \leq 1$, a pathogen cannot invade a fully susceptible host population. When $R_0 > 1$, a pathogen can invade a fully susceptible host population with probability $1 - (1/R_0)$ (Gerber et al., 2005; Allen, 2015).

In Section 5.B we show that, consistent with continuous-time disease models (Rohani et al., 2009), R_0 for discrete-time IPMs with an environmental reservoir is composed of transmission from host contact and the environment. We use this result in combination with our fully parameterized host-parasite IPM to calculate the temperature-dependent R_0 for *R. muscosa*-Bd systems with and without infection from the environmental zoospore pool.

5.3.3 The hybrid model

The model described by equations 1-3 is sufficient to describe the dynamics of an initial epizootic, but to examine Bd-induced extinction dynamics in *R. muscosa* populations a number of additions need to be made. We briefly describe the hybrid model that accounts for the within-year Bd dynamics as well as the between year demography of *R. muscosa*. Fig. 5.1 gives a visual representation of the hybrid model and Section 5.C gives a full description.

The within-year component (Fig. 5.1B.), is identical to the IPM given in equations 1-3 with the addition of three tadpole stages. The tadpole stage of *R. muscosa* is likely important in generating enzootic dynamics in *R. muscosa* populations (Briggs et al., 2005, 2010). We assumed all tadpoles were immediately infected with Bd and had a constant mean contribution to the zoospore pool (Table 1). This is justified by the observation that most tadpoles in *R. muscosa* populations carry high fungal loads, even in enzootic populations (Briggs et al., 2010). *R. muscosa* tadpole survival is not affected by Bd infection. Therefore, the within-season dynamics of the tadpole stages were simply given by the probability of a tadpole surviving from time t to $t+1$. Infected tadpoles also contributed to the zoospore pool at each time step (Fig. 5.1B).

R. muscosa populations also experience seasonal temperature fluctuations in which winter lake temperatures drop to approximately 4 °C in the winter (in the unfrozen portion of a lake where the frogs overwinter) and reach approximately 20 °C in the summer (Knapp et al., 2011). We accounted for this seasonal variability by imposing a deterministically fluctuating environment on the *R. muscosa*-Bd IPM (Fig. 5.1A). At each discrete time point within a season, a new temperature was calculated and the temperature-dependent vital rate functions were updated accordingly.

The between-year component of the hybrid model accounted for yearly maturation and metamorphosis of the tadpole stages as well as reproduction of adults (Fig. 5.1C). We assumed that the recruitment of metamorphed tadpoles into the adult stage was density-dependent and that all tadpoles entered the adult stage as uninfected (i.e. all individuals are infected as tadpoles, but lose their infection at metamorphosis, Briggs et al., 2010). Because we have no empirical evidence for Bd-induced fertility reduction in *R. muscosa*, we assumed that reproduction in uninfected and infected adults was the same.

5.3.4 Simulating the hybrid model

After parameterizing the hybrid model using the above experiments, we used the model to make predictions about the probability of disease-induced host extinctions. Because demographic stochasticity is important when predicting extinction for small populations (Lande et al., 2003), we included it into the hybrid model (Caswell, 2001; Schreiber and Ross, 2016). To do this, we first discretized the within-season IPM using the mid-point rule and 30 mesh points (Easterling et al., 2000) and then determined the transition of an individual frog or zoospore to another state (including death) as a draw from a multinomial distribution with probabilities given by the discretized hybrid model at that time step (Section 5.D). In addition, we assumed that both the production of tadpoles that occurs once a year in the spring and the number of zoospores shed into the zoospore pool at each time step followed a Poisson distribution (Section 5.D).

To answer our question regarding the importance of transmission, resistance, and tolerance on Bd-induced extinction, we performed two analyses. First, we examined how different transmission functions parameterized from our mesocosm experiment affected extinction risk. Using the three transmission functions with a dynamic zoospore pool described in Table 2 and the parameter values given in Table 1, we performed 500 stochastic simulations of the hybrid model to generate time-dependent extinction curves over a 25 year period. All simulations were started with 10 uninfected adult frogs, 85 year-one tadpoles (T_1), 12 year-two tadpoles (T_2), and 3 year-three tadpoles (T_3). The relative proportions of adult frogs and tadpoles were assigned based on the stable stage distribution in the Bd-free model. While the initial conditions necessarily affect the absolute time to extinction, they do not affect the shapes of the extinction curves for different transmission functions relative to each other. All simulations started in the winter at 4 °C, with reproduction occurring in the first spring at 12 °C (Fig. 5.1A). Given our analysis of R_0 in the previous section, we assumed that Bd could invade with a probability of one, such that tadpoles were immediately infected with Bd and began contributing to the zoospore pool $Z(t)$ (see Section 5.D). For each simulation, we calculated time-dependent extinction curves as the mean probability of going extinct in a given time step over all 500 simulations.

Second, we performed a sensitivity analysis on the hybrid model to assess the relative importance of transmission compared to resistance and tolerance. Transmission was determined by the parameters in the transmission function ϕ, resistance was determined by the parameters in the growth function

$G(x', x)$, the loss of infection function $l(x)$, and the initial infection burden function $G_0(x')$, and tolerance was determined by the parameters in the survival function $s(x)$. To perform the sensitivity analysis, we ran 1000 simulations using the parameter values given in Table 1 and the initial conditions described above. For each simulation we recorded whether or not a *R. muscosa* population went extinct in $\leq$ 8 years, as this was where extinction probability was approximately 50% with the default parameter values.

On each run of the simulation we perturbed sixteen lower-level transmission, resistance and tolerance parameters by, for each parameter, drawing a random number from a log normal distribution with median 1 and dispersion parameter $\sigma_{\text{sensitivity}} = 0.3$ and multiplying the given parameter by this random number (Sobie, 2009). Our results were robust to our choice of σ and our method of perturbation (Fig. 5.S3). For each simulation, we saved the perturbed parameter values and stored them in a 1000 by 16 parameter matrix. Upon completion of the simulation, we used both regularized logistic regression and a Random Forest classifier in which our response variable was whether or not a given simulation went extinct and our predictor variables were the scaled (i.e. z-transformed) matrix of perturbed predictors (Harper et al., 2011; Pedregosa et al., 2011; Lee et al., 2013). Using this approach, we could then identify the relative importance of each parameter in the vital rate functions in predicting whether or not extinction occurred (Sobie, 2009). Moreover, we also built a pruned regression tree to visualize the interactive effects of transmission, resistance, and tolerance parameters on host extinction risk (Harper et al., 2011). All the code necessary to replicate the analyses is provided at `https://github.com/mqwilber/host_extinction`.

5.4 Results

5.4.1 Question 1: What is the structure of the transmission function?

Accounting for the dynamics of the environmental zoospore pool resulted in significantly better transmission models compared to those that did not. The transmission model with density-dependent host to host transmission as well as transmission from a dynamic zoospore pool was a better model

than all other transmission models considered (Table 2). In addition to being a better model in terms of WAIC, the density-dependent transmission model also captured the marginal pattern of increasing probability of infection with increasing density of infected hosts (Fig. 5.S4).

5.4.2 Question 2: How does the transmission function affect the ability of Bd to invade?

Using the temperature-dependent vital functions described in Table 1 and the best-fitting density-dependent transmission function with an environmental zoospore pool (Table 2), we examined how host density and temperature affected the ability of Bd to invade a *R. muscosa* population. When transmission was density-dependent, but did not depend on the environmental zoospore pool, Bd was able to invade *R. muscosa* populations over a large range of densities and temperatures, though there was a slight protective effect of low temperatures and low densities (Fig. 5.2A). Including transmission from the environmental zoospore pool substantially increased the region in which Bd could invade and invasion was highly probable for most temperatures and host densities (Fig. 5.2B, C).

5.4.3 Question 3: How sensitive is disease-induced extinction to transmission, resistance, and tolerance?

The time-dependent probability of extinction was similar between the three transmission functions that included a dynamic zoospore pool (Fig. 5.S5, Table 2). The similarity between these curves was due to the overwhelming influence of the infection probability from the zoospore pool, which swamped out the well-known differences between frequency-dependent and density-dependent transmission functions. Drastically decreasing zoospore survival probability below what has been observed in laboratory experiments (Woodhams et al., 2008), led to an expected reduction in extinction risk as the transmission probability then declined with decreasing host density (given density-dependent transmission, Fig. 5.S5).

A sensitivity analysis of Bd-induced host extinction to transmission, host resistance, and host tolerance showed that, regardless of the transmission function used, *R. muscosa* extinction was more sensitive to the parameters relating to host resistance and tolerance than to parameters relating to

transmission (Fig. 5.3). In particular, the most important parameter across all transmission functions was the slope of the growth function $b_{1,1}$, which is a parameter affecting host resistance. For a given temperature, the logistic regression analysis showed that decreasing this parameter, which roughly corresponds to decreasing the mean Bd load on a host for a given temperature, decreased the probability of disease-induced extinction for all transmission functions (Fig. 5.3A-C). Bd-induced *R. muscosa* extinction was also sensitive to the parameters of the survival function, particularly the intercept of the survival function $b_{0,0}$. This parameter can be thought of as the threshold at which Bd-induced mortality begins to occur given a fixed slope in the survival function. The logistic regression analysis showed that increasing this parameter, which corresponds to increasing the threshold at which *R. muscosa* begins to suffer load-dependent Bd mortality, decreased the probability of extinction (Fig. 5.3A-C).

Resistance and tolerance parameters also showed significant interactions when affecting host extinction risk. Random forests and pruned regression trees showed the importance of the slope of the growth function as well as the importance of the interaction between this parameter and the intercept of the survival function (a tolerance parameter) and the temperature-dependency in the growth function (a resistance parameter) (Fig. 5.3A-C; Fig. 5.S6-8).

5.5 Discussion

Wildlife conservation in the face of disease emphasizes the importance of the transmission function in extinction risk (McCallum, 2012). This is a reasonable emphasis as the transmission function is ultimately the most important aspect of disease-induced extinction: if a host does not get infected with a disease it will not suffer disease-induced mortality. In amphibian-Bd systems it has been hypothesized that both amphibian density and an environmental pool of zoospores can affect transmission (Rachowicz and Briggs, 2007; Briggs et al., 2010; Courtois et al., 2017), but the nature of this transmission has rarely been quantified. We experimentally quantified the transmission function in the *R. muscosa*-Bd system and used these results, in combination with a dynamic model, to predict how the environmental zoospore reservoir affected the ability of Bd to invade an amphibian population. Consistent with previous theory (Godfray et al., 1999; Rohani et al., 2009), we found that including an environmental zoospore pool substantially increased R_0 for *R. muscosa*-Bd systems, such that Bd was able to invade a *R. muscosa* population for most realistic temperatures and host densities. To our knowledge, this is

the first estimation of R_0 in an amphibian-Bd system (but see Woodhams et al., 2011, for a discussion of R_0 in amphibian-Bd systems), and the large value of R_0 across all temperatures and densities is consistent with field observations that temperature and density have little protective effect in the *R. muscosa/sierrae* system (Knapp et al., 2011, R. A. Knapp et al., unpublished). These results suggest that attempting to prevent Bd invasion into a system may be largely futile and management should be focused on mitigating post-invasion Bd impacts (Langwig et al., 2015).

Conditional on Bd invasion, we used our parameterized model to explore the importance of the transmission function on Bd-induced amphibian extinction and found that the extinction dynamics were similar between all transmission models with a dynamic zoospore pool. This was due to the large number of zoospores shed by infected amphibians combined with the laboratory-estimated decay rate of zoospores outside the host, leading to a zoospore pool that remained large even for rapidly declining host populations (Fig. 5.1D). Only considering this result, we would then expect *R. muscosa* populations to be at substantial risk of disease-induced extinction given that the persisting zoospore pool prevents a decrease in transmission rate with declining host density (Anderson and May, 1981; Godfray et al., 1999; De Castro and Bolker, 2005). This finding is consistent with a number of other studies that have found that the dynamics of the Bd zoospore pool are critical for determining Bd-induced amphibian extinctions (Mitchell et al., 2008; Briggs et al., 2010; Doddington et al., 2013). If abiotic or biotic factors such as temperature, stream flow, water chemistry, and/or zoospore consumption by aquatic organisms are able to substantially increase zoospore death rate beyond the values seen in the lab (Tunstall, 2012; Strauss and Smith, 2013; Venesky et al., 2013; Heard et al., 2014; Schmeller et al., 2014), then we might expect a reduction, though not an elimination, of Bd invasion probability and amphibian extinction risk.

However, considering only the transmission function ignores the fact that, conditional on infection, increasing resistance or tolerance to a disease can also reduce disease-induced mortality and thus provide alternative mechanisms by which to manage disease-induced extinction risk (Kilpatrick, 2006; Vander Wal et al., 2014; Langwig et al., 2015). Using our model, we found that Bd-induced extinction risk was far more sensitive to host resistance and tolerance than to the transmission dynamics of Bd. In particular, extinction risk of *R. muscosa* populations was most sensitive to the vital rate functions dictating the growth rate of Bd on a host, the load-dependent survival probability, and the interaction between these two functions.

This result highlights the importance of accounting for the load-dependent nature of vital rate functions when considering extinction risk in load-dependent diseases such as chytridiomycosis. In this study, consistent with results observed in the field (Vredenburg et al., 2010), the survival function of Bd was strongly non-linear such that above ≈ 9 ln zoospores $= 8103$ zoospores the survival probability of *R. muscosa* rapidly declined (Fig. 5.S1B). When the survival function is load-dependent and highly non-linear, as observed in some amphibian-Bd systems (Stockwell et al., 2010; Vredenburg et al., 2010, but see Clare et al. (2016)), small changes in host resistance or tolerance can lead to abrupt changes in survival probability. In particular, non-linearities in the survival function (i.e. tolerance) need to be considered in the context of the shape of the growth function (i.e. resistance) of a parasite on its host.

To illustrate the generality of this result, consider the following graphical argument. Take a pathogen growth function ($G(x', x)$) that predicts a static mean pathogen load near the threshold at which the survival function predicts a drastic decrease in survival probability (Fig. 5.4, i.e. a non-linear dose-response curve; Dwyer et al., 1997; Handel and Rohani, 2015; Louie et al., 2016). Slightly shifting the slope (or the intercept) of the growth function (i.e. changing resistance) up or down will increase or decrease the static mean pathogen load and move a host into the region of the survival curve where either mortality or survival is almost certain (Fig. 5.4). Similarly, holding the growth function constant and altering the survival function (i.e. changing tolerance) will change how close the static mean pathogen load is to the survival function threshold (Fig. 5.4). This suggests that identifying how the growth function of a pathogen changes in resistant host populations, whether by decreasing the slope, decreasing the intercept or by transitioning from a linear to a non-linear function (Langwig et al., 2017), is important for understanding the sensitivity of extinction risk to both the growth function (resistance) and the survival function (tolerance). Identifying whether or not a disease system shows this strong interaction between resistance and tolerance can help determine whether disease mitigation should focus on reducing parasite loads via strategies such as inducing acquired immunity, microbial treatments, or selecting for resistance (Harris et al., 2009; Langwig et al., 2015) or decreasing parasite transmission via strategies such as culling and/or treating the disease reservoir (Cleaveland et al., 2001).

The importance of host resistance and tolerance for our model's predictions of disease-induced extinction indicates that these host strategies could promote population persistence in *R. muscosa*-Bd populations, as they have in other species of amphibians suffering from chytridiomycosis, bats suffering from white-nose syndrome, and Tasmanian devils suffering from facial tumor disease (Hoyt et al.,

2016; Savage and Zamudio, 2016; Epstein et al., 2016; Langwig et al., 2017). In fact, a recent study showed that many populations of *R. sierrae* that experienced Bd-induced population declines over the last four decades are recovering in the presence of Bd, but with reduced Bd loads relative to naive populations (Knapp et al., 2016). This result is consistent with resistance mechanisms reducing Bd load and thus increasing host survival probability. While changes in resistance mechanisms, and not tolerance mechanisms, are putatively responsible for persistence in a number of load-dependent diseases (Savage and Zamudio, 2011; Epstein et al., 2016; Langwig et al., 2017), our results show that the shape of the tolerance function is critical for understanding the effects of resistance on host persistence (Fig. 5.4). An important future direction will be to explore how heterogeneities in host resistance, tolerance, and/or transmission functions can promote host persistence (Boots et al., 2009; Langwig et al., 2015; Brunner et al., 2017).

While our study suggests that an interaction between resistance and tolerance promotes population persistence in *R. musocsa*-Bd populations, additional mechanisms that we do not consider here are likely important in other amphibian-Bd systems. For example, laboratory studies have shown that Bd can evolve increased or reduced virulence in as few as 50 generations (less than one year, Langhammer et al., 2013; Voyles et al., 2014; Refsnider et al., 2015). If Bd virulence attenuates over the course of an epizootic, these changes could augment host persistence, without any changes in host resistance or tolerance. In reality, the evolution of both the pathogen and the host affects population persistence (Vander Wal et al., 2014) and developing load-dependent models that incorporate both of these processes is an open challenge in wildlife epidemiology.

The emergence of a number of diseases of conservation concern have highlighted the importance of considering disease-induced extinction in the context of load-dependent parasite dynamics. We show that simultaneously considering transmission, resistance, and tolerance, in conjunction with load-dependent dynamics, provides novel insight regarding how to best manage emerging infectious diseases.

5.6 Acknowledgments

We would like to acknowledge the Sierra Nevada Aquatic Research Laboratory for logistical support for the experiments described in this work. We also acknowledge W. Murdoch, R. Nisbet, and three anonymous reviewers for helpful feedback on this manuscript and S. Weinstein for her artis-

tic contribution. Funding was provided by the National Science Foundation, USA (NSF EEID grant DEB-0723563 and NSF LTREB grant DEB-1557190). M. Wilber was supported by an NSF Graduate Research Fellowship (Grant No. DGE 1144085) and the University of California Regents (USA).

References

Allen, L. J. S. 2015. Stochastic Population and Epidemic Models: Persistence and Extinction. Springer International Publishing, London.

Anderson, R. M., and R. M. May. 1981. The population dynamics of microparasites and their invertebrate hosts. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **291**:451–524.

Andre, S. E., J. Parker, and C. J. Briggs. 2008. Effect of temperature on host response to *Batrachochytrium dendrobatidis* infection in the mountain yellow-legged frog (*Rana muscosa*). Journal of Wildlife Diseases **44**:716–720.

Bielby, J., M. C. Fisher, F. C. Clare, G. M. Rosa, and T. W. J. Garner. 2015. Host species vary in infection probability, sub-lethal effects, and costs of immune response when exposed to an amphibian parasite. Scientific Reports **5**:1–8.

Blehert, D. S., A. C. Hicks, M. Behr, C. U. Meteyer, B. M. Berlowski-Zier, E. L. Buckles, J. T. H. Coleman, S. R. Darling, A. Gargas, R. Niver, J. C. Okoniewski, R. J. Rudd, and B. Ward. 2008. Bat white-nose syndrome : an emerging fungal pathogen? Science **323**:227.

Boots, M., A. Best, M. R. Miller, and A. White. 2009. The role of ecological feedbacks in the evolution of host defence: what does theory tell us? Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **364**:27–36.

Boyle, D. G., D. B. Boyle, V. Olsen, J. A. T. Morgan, and A. D. Hyatt. 2004. Rapid quantitative detection of chytridiomycosis (*Batrachochytrium dendrobatidis*) in amphibian samples using real-time Taqman PCR assay. Diseases of Aquatic Organisms **60**:141–8.

Briggs, C. J., R. A. Knapp, and V. T. Vredenburg. 2010. Enzootic and epizootic dynamics of the chytrid fungal pathogen of amphibians. Proceedings of the National Academy of Sciences of the United States of America **107**:9695–9700.

Briggs, C. J., V. T. Vredenburg, R. A. Knapp, and L. J. Rachowicz. 2005. Investigating the population-level effects of chytridiomycosis: an emerging infectious disease of amphibians. Ecology **86**:3149–3159.

Brunner, J. L., L. Beaty, A. Guitard, and D. Russell. 2017. Heterogeneities in the infection process drive ranavirus transmission. Ecology **98**:576–582.

Caswell, H. 2001. Matrix Population Models: Construction, Analysis, and Interpretation. Second edition. Sinauer, Sunderland, Massachusetts.

Clare, F., O. Daniel, T. Garner, and M. Fisher. 2016. Assessing the ability of swab data to determine the true burden of infection for the amphibian pathogen *Batrachochytrium dendrobatidis*. EcoHealth **13**:360–367.

Cleaveland, S., G. R. Hess, A. P. Dobson, M. K. Laurenson, H. I. McCallum, M. G. Roberts, and R. Woodroffe, 2001. The role of pathogens in biological conservation. Pages 139–150 *in* P. J. Hudson, A. Rizzoli, B. Grenfell, H. Heesterbeek, and A. Dobson, editors. The Ecology of Wildlife Diseases. Oxford University Press, Oxford.

Courtois, E. A., A. Loyau, M. Bourgoin, and D. S. Schmeller. 2017. Initiation of *Batrachochytrium dendrobatidis* infection in the absence of physical contact with infected hosts - a field study in a high altitude lake. Oikos **126**:843–851.

Daszak, P., A. A. Cunningham, and A. D. Hyatt. 2003. Infectious disease and amphibian population declines. Diversity and Distributions **9**:141–150.

De Castro, F., and B. Bolker. 2005. Mechanisms of disease-induced extinction. Ecology Letters **8**:117–126.

Diekmann, O., J. A. P. Heesterbeek, and J. A. J. Metz. 1990. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology **28**:365–382.

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

Doddington, B. J., J. Bosch, J. A. Oliver, N. C. Grassly, G. Garcia, B. R. Schmidt, T. W. J. Garner, and M. C. Fisher. 2013. Context-dependent amphibian host population response to an invading pathogen. Ecology **94**:1795–1804.

Dwyer, G., J. S. Elkinton, and J. P. Buonaccorsi. 1997. Host heterogeneity in susceptibility and disease dynamics: tests of a mathematical model. The American Naturalist **150**:685–707.

Easterling, M. R., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new structured population model. Ecology **81**:694–708.

Ellison, A. R., T. Tunstall, G. V. Direnzo, M. C. Hughey, E. A. Rebollar, L. K. Belden, R. N. Harris, R. Ibáñez, K. R. Lips, and K. R. Zamudio. 2015. More than skin deep: Functional genomic basis for resistance to amphibian chytridiomycosis. Genome Biology and Evolution **7**:286–298.

Epstein, B., M. Jones, R. Hamede, S. Hendricks, H. McCallum, E. P. Murchison, B. Schonfeld, C. Wiench, P. Hohenlohe, and A. Storfer. 2016. Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications **7**:12684.

Forrest, M. J., and M. A. Schlaepfer. 2011. Nothing a hot bath won't cure: infection rates of amphibian chytrid fungus correlate negatively with water temperature under natural field settings. PLoS ONE **6**:e28444.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. Bayesian Data Analysis. 3 edition. Taylor & Francis Group, LLC, Boca Raton.

Gerber, L. R., H. McCallum, K. D. Lafferty, J. L. Sabo, and A. Dobson. 2005. Exposing extinction risk analysis to pathogens: Is disease just another form of density dependence? Ecological Applications **15**:1402–1414.

Godfray, H. C. J., C. J. Briggs, N. D. Barlow, M. O'Callaghan, T. R. Glare, and T. A. Jackson. 1999. A model of insect-pathogen dynamics in which a pathogenic bacterium can also reproduce saprophytically. Proceedings of the Royal Society B: Biological Sciences **266**:233–240.

Grimmett, G., and D. Stirzaker. 2001. Probability and Random Processes. Oxford University Press, Oxford, United Kingdom.

Grogan, L. F., A. D. Phillott, B. C. Scheele, L. Berger, S. D. Cashins, S. C. Bell, R. Puschendorf, and L. F. Skerratt. 2016. Endemicity of chytridiomycosis features pathogen overdispersion. Journal of Animal Ecology **85**:806–816.

Handel, A., and P. Rohani. 2015. Crossing the scale from within-host infection dynamics to between-host transmission fitness: a discussion of current assumptions and knowledge. Philosophical Transactions of the Royal Society B: Biological Sciences **370**:20140302.

Harper, E. B., J. C. Stella, and A. K. Fremier. 2011. Global sensitivity analysis for complex ecological models: a case study of riparian cottonwood population dynamics. Ecological Applications **21**:1225–1240.

Harris, R. N., R. M. Brucker, J. B. Walke, M. H. Becker, C. R. Schwantes, D. C. Flaherty, B. a. Lam, D. C. Woodhams, C. J. Briggs, V. T. Vredenburg, and K. P. C. Minbiole. 2009. Skin microbes on frogs prevent morbidity and mortality caused by a lethal skin fungus. The ISME Journal **3**:818–824.

Heard, G. W., M. P. Scroggie, N. Clemann, and D. S. L. Ramsey. 2014. Wetland characteristics influence disease risk for a threatened amphibian Wetland characteristics influence disease risk for a threatened amphibian. Ecological Applications **24**:650–662.

Hoyt, J. R., K. E. Langwig, K. Sun, G. Lu, K. L. Parise, T. Jiang, W. F. Frick, J. T. Foster, J. Feng, and A. M. Kilpatrick. 2016. Host persistence or extinction from emerging infectious disease: insights from white-nose syndrome in endemic and invading regions. Proceedings of the Royal Society B: Biological Sciences **283**:20152861.

Jani, A. J., and C. J. Briggs. 2014. The pathogen *Batrachochytrium dendrobatidis* disturbs the frog skin microbiome during a natural epidemic and experimental infection. Proceedings of the National Academy of Sciences **111**:E5049–E5058.

Jenkinson, T. S., C. M. Betancourt Román, C. Lambertini, A. Valencia-Aguilar, D. Rodriguez, C. H. L. Nunes-De-Almeida, J. Ruggeri, A. M. Belasen, D. Da Silva Leite, K. R. Zamudio, J. E. Longcore,

L. F. Toledo, and T. Y. James. 2016. Amphibian-killing chytrid in Brazil comprises both locally endemic and globally expanding populations. Molecular Ecology **25**:2978–2996.

Kilpatrick, A. M. 2006. Facilitating the evolution of resistance to avian malaria in Hawaiian birds. Biological Conservation **128**:475–485.

Kilpatrick, A. M., C. J. Briggs, and P. Daszak. 2010. The ecology and impact of chytridiomycosis: an emerging disease of amphibians. Trends in Ecology and Evolution **25**:109–118.

Klepac, P., and H. Caswell. 2011. The stage-structured epidemic: Linking disease and demography with a multi-state matrix approach model. Theoretical Ecology **4**:301–319.

Knapp, R. A., C. J. Briggs, T. C. Smith, and J. R. Maurer. 2011. Nowhere to hide: impact of a temperature-sensitive amphibian pathogen along an elevation gradient in the temperate zone. Ecosphere **2**:art93.

Knapp, R. A., G. M. Fellers, P. M. Kleeman, D. A. W. Miller, V. T. Vredenburg, E. B. Rosenblum, and C. J. Briggs. 2016. Large-scale recovery of an endangered amphibian despite ongoing exposure to multiple stressors. Proceedings of the National Academy of Sciences **113**:11889–11894.

Lande, R., S. Engen, and B.-E. Saether. 2003. Stochastic Population Dynamics in Ecology and Conservation. Oxford University Press, Oxford.

Langhammer, P. F., K. R. Lips, P. A. Burrowes, T. Tunstall, C. M. Palmer, and J. P. Collins. 2013. A fungal pathogen of amphibians, Batrachochytrium dendrobatidis, attenuates in pathogenicity with in vitro passages. PLoS ONE **8**:1–9.

Langwig, K. E., J. Hoyt, K. Parise, W. Frick, J. Foster, and A. M. Kilpatrick. 2017. Resistance in persisting bat populations after white-nose syndrome invasion. Philosophical Transactions of the Royal Society B **372**:20160044.

Langwig, K. E., J. Voyles, M. Q. Wilber, W. F. Frick, K. A. Murray, B. M. Bolker, J. P. Collins, T. L. Cheng, M. C. Fisher, J. R. Hoyt, D. L. Lindner, H. I. McCallum, R. Puschendorf, E. B. Rosenblum, M. Toothman, C. K. Willis, C. J. Briggs, and A. M. Kilpatrick. 2015. Context-dependent conservation responses to emerging wildlife diseases. Frontiers in Ecology and the Environment **13**:195–202.

Lee, Y.-S., O. Z. Liu, H. S. Hwang, B. C. Knollmann, and E. A. Sobie. 2013. Parameter sensitivity analysis of stochastic models provides insights into cardiac calcium sparks. Biophysical Journal **104**:1142–1150.

Louie, A., K. H. Song, A. Hotson, A. Thomas Tate, and D. S. Schneider. 2016. How many parameters does it take to describe disease tolerance? PLoS Biology **14**:e1002435.

McCallum, H. 2012. Disease and the dynamics of extinction. Philosophical Transactions of the Royal Society B **367**:2828–39.

McCallum, H., N. Barlow, and J. Hone. 2001. How should pathogen transmission be modelled? Trends in Ecology and Evolution **16**:295–300.

McCallum, H., and A. Dobson. 1995. Detecting disease and parasite threats to endangered species and ecosystems. Trends in Ecology and Evolution **10**:190–194.

McCallum, H., M. Jones, C. Hawkins, R. Hamede, S. Lachish, D. L. Sinn, N. Beeton, and B. Lazenby. 2009. Transmission dynamics of Tasmanian devil facial tumor disease may lead to disease-induced extinction. Ecology **90**:3379–3392.

Medzhitov, R., D. S. Schneider, and M. P. Soares. 2012. Disease tolerance as a defense strategy. Science **335**:936–941.

Metcalf, C. J. E., A. L. Graham, M. Martinez-Bakker, and D. Z. Childs. 2015. Opportunities and challenges of Integral Projection Models for modelling host-parasite dynamics. Journal of Animal Ecology **85**:343–355.

Mitchell, K. M., T. S. Churcher, T. W. J. Garner, and M. C. Fisher. 2008. Persistence of the emerging pathogen *Batrachochytrium dendrobatidis* outside the amphibian host greatly increases the probability of host extinction. Proceedings of the Royal Society B **275**:329–334.

Oli, M. K., M. Venkataraman, P. A. Klein, L. D. Wendland, and M. B. Brown. 2006. Population dynamics of infectious diseases: A discrete time model. Ecological Modelling **198**:183–194.

Orr, H. A., and R. L. Unckless. 2014. The population genetics of evolutionary rescue. PLoS Genetics **10**:e1004551.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research **12**:2825–2830.

Piotrowski, J. S., S. L. Annis, and J. E. Longcore. 2004. Physiology of *Batrachochytrium dendrobatidis*, a chytrid pathogen of amphibians. Mycologia **96**:9–15.

Rachowicz, L. J., and C. J. Briggs. 2007. Quantifying the disease transmission function: effects of density on *Batrachochytrium dendrobatidis* transmission in the mountain yellow-legged frog *Rana muscosa*. The Journal of Animal Ecology **76**:711–21.

Refsnider, J. M., T. J. Poorten, P. F. Langhammer, P. A. Burrowes, and E. B. Rosenblum. 2015. Genomic correlates of virulence attenuation in the deadly amphibian chytrid fungus, Batrachochytrium dendrobatidis. G3: Genes—Genomes—Genetics **5**:2291–2298.

Retallick, R. W. R., H. McCallum, and R. Speare. 2004. Endemic infection of the amphibian chytrid fungus in a frog community post-decline. PLoS Biology **2**:e351.

Rohani, P., R. Breban, D. E. Stallknecht, and J. M. Drake. 2009. Environmental transmission of low pathogenicity avian influenza viruses and its implications for pathogen invasion. Proceedings of the National Academy of Sciences **106**:10365–10369.

Rosenblum, E. B., T. Y. James, K. R. Zamudio, T. J. Poorten, D. Ilut, D. Rodriguez, J. M. Eastman, K. Richards-Hrdlicka, S. Joneson, T. S. Jenkinson, J. E. Longcore, G. Parra Olea, L. F. Toledo, M. L. Arellano, E. M. Medina, S. Restrepo, S. V. Flechas, L. Berger, C. J. Briggs, and J. E. Stajich. 2013. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. Proceedings of the National Academy of Sciences **110**:9385–90.

Roy, B. A., and J. W. Kirchner. 2000. Evolutionary eynamics of pathogen resistance and tolerance. Evolution **54**:51–63.

Savage, A. E., and K. R. Zamudio. 2011. MHC genotypes associate with resistance to a frog-killing fungus. Proceedings of the National Academy of Sciences **108**:16705–16710.

Savage, A. E., and K. R. Zamudio. 2016. Adaptive tolerance to a pathogenic fungus drives major histocompatibility complex evolution in natural amphibian populations. Proceedings of the Royal Society B **283**:20153115.

Schmeller, D. S., M. Blooi, A. Martel, T. W. J. Garner, M. C. Fisher, F. Azemar, F. C. Clare, C. Leclerc, L. Jäger, M. Guevara-Nieto, A. Loyau, and F. Pasmans. 2014. Microscopic aquatic predators strongly affect infection dynamics of a globally emerged pathogen. Current Biology **24**:176–180.

Schreiber, S. J., and N. Ross. 2016. Individual-based integral projection models: the role of size-structure on extinction risk and establishment success. Methods in Ecology and Evolution **7**:867–874.

Skerratt, L. F., L. Berger, R. Speare, S. Cashins, K. R. McDonald, A. D. Phillott, H. B. Hines, and N. Kenyon. 2007. Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs. EcoHealth **4**:125–134.

Smith, K. F., D. F. Sax, and K. D. Lafferty. 2006. Evidence for the role of infectious disease in species extinction and endangerment. Conservation Biology **20**:1349–1357.

Sobie, E. A. 2009. Parameter sensitivity analysis in electrophysiological models using multivariable regression. Biophysical Journal **96**:1264–1274.

Stockwell, M. P., J. Clulow, and M. J. Mahony. 2010. Host species determines whether infection load increases beyond disease-causing thresholds following exposure to the amphibian chytrid fungus. Animal Conservation **13**:62–71.

Strauss, A., and K. G. Smith. 2013. Why does amphibian chytrid (*Batrachochytrium dendrobatidis*) not occur everywhere? An exploratory study in missouri ponds. PLoS ONE **8**:e76035.

Tunstall, T. S., 2012. Characteristics of the emergent disease *Batrachochytrium dendrobatidis* in *Rana muscosa* and *Rana sierrae* species complex. Ph.D. thesis, University of California, Berkeley.

Vander Wal, E., D. Garant, S. Calmé, C. A. Chapman, M. Festa-Bianchet, V. Millien, S. Rioux-Paquette, and F. Pelletier. 2014. Applying evolutionary concepts to wildlife disease ecology and management. Evolutionary Applications **7**:856–868.

Venesky, M. D., X. Liu, E. L. Sauer, and J. R. Rohr. 2013. Linking manipulative experiments to field data to test the dilution effect. Journal of Animal Ecology **83**:557–565.

Voyles, J., L. Berger, S. Young, R. Speare, R. Webb, J. Warner, D. Rudd, R. Campbell, and L. Skerratt. 2007. Electrolyte depletion and osmotic imbalance in amphibians with chytridiomycosis. Diseases of Aquatic Organisms **77**:113–118.

Voyles, J., L. R. Johnson, C. J. Briggs, S. D. Cashins, R. A. Alford, L. Berger, L. F. Skerratt, R. Speare, and E. B. Rosenblum. 2014. Experimental evolution alters the rate and temporal pattern of population growth in Batrachochytrium dendrobatidis, a lethal fungal pathogen of amphibians. Ecology and Evolution **4**:3633–3641.

Voyles, J., S. Young, L. Berger, C. Campbell, W. F. Voyles, and A. Dinudom. 2009. Pathogenesis of chytridiomycosis, a cause of catastrophic amphibian declines. Science **326**:5–8.

Vredenburg, V. T., R. A. Knapp, T. S. Tunstall, and C. J. Briggs. 2010. Dynamics of an emerging disease drive large-scale amphibian population extinctions. Proceedings of the National Academy of Sciences **107**:9689–94.

Wilber, M. Q., K. E. Langwig, A. M. Kilpatrick, H. I. Mccallum, and C. J. Briggs. 2016. Integral Projection Models for host-parasite systems with an application to amphibian chytrid fungus. Methods in Ecology and Evolution **7**:1182–1194.

Woodhams, D. C., R. A. Alford, C. J. Briggs, M. Johnson, and L. A. Rollins-Smith. 2008. Life-history trade-offs influence disease in changing climates: Strategies of an amphibian pathogen. Ecology **89**:1627–1639.

Woodhams, D. C., J. Bosch, C. J. Briggs, S. Cashins, L. R. Davis, A. Lauer, E. Muths, R. Puschendorf, B. R. Schmidt, B. Sheafor, and J. Voyles. 2011. Mitigating amphibian disease: strategies to maintain wild populations and control chytridiomycosis. Frontiers in Zoology **8**:8.

Table 5.1: Parameters used in the *Rana muscosa*-Bd hybrid model. All b and c parameters had a Normal prior with mean 0 and standard deviation 5. All ν parameters had a half-Cauchy prior from 0 to ∞ with scale parameter 1. For all statistical models, convergence was assessed using traceplots and ensuring that the Gelman-Rubin statistic was < 1.05 (Gelman et al., 2014). *logit* specifies a logistic link, x is ln zoospore load, and T is temperature. All probabilities are over a three day time step.

Description	**Functional Form**	**Parameters**	**Details of Parameterization**
Infected survival function, $s(x)$: Probability of host survival from t to $t+1$ with load x	$\text{logit}[s(x)] = b_{0,0} + b_{1,0}x$	$b_{0,0} = 5.295$ $b_{1,0} = -2.595$	*Likelihood* Bernoulli($s(x)$), x was z-transformed Fig. 5.S1B.
Growth function, $G(x', x)$: Probability density of host transitioning from load x to load x' at time $t+1$	$\mu(x,T) = b_{0,1} + b_{1,1}x + b_{2,1}T$ $\sigma^2(x) = \nu_{0,1}\exp(2c_{0,1}x)$	$b_{0,1} = 0.012$ $b_{1,1} = 0.799$ $b_{2,1} = 0.092$ $\nu_{0,1} = 5.92$ $c_{0,1} = -0.049$	*Likelihood* Normal($\mu(x,T), \sigma^2(x)$) Fig. 5.S1A.
Loss of infection function, $l(x)$: The probability of host having infection of load x at t and losing it by $t+1$	$\text{logit}[l(x,T)] = b_{0,2} + b_{1,2}x + b_{2,2}T$	$b_{0,2} = 1.213$ $b_{1,2} = -0.472$ $b_{2,2} = -0.151$	*Likelihood* Bernoulli($l(x,T)$) Fig. 5.S1C.
Initial infection burden function, $G_0(x')$: Probability density of being uninfected at t and gaining an infection of x' at $t+1$	$\mu(T) = b_{0,3} + b_{1,3}T$ $\sigma^2(T) = \nu_{0,3}\exp(2c_{0,3}T)$	$b_{0,3} = 0.642$ $b_{1,3} = 0.137$ $\nu_{0,3} = 0.59$ $c_{0,3} = 0.063$	*Likelihood* Normal($\mu(T), \sigma^2(T)$) Fig. 5.S1D.
Transmission function, ϕ: The probability of an uninfected host gaining an infection at time $t+1$	Functional forms vary	See Table 2	See Table 2
Uninfected adult survival probability over three day time step, s_0	Constant	$s_0 = 0.999$	Yearly survival from Briggs *et al.* 2005 converted to a three day time scale
Temperature-dependent zoospore survival probability, ν	$\nu = f(T)$	Non-parametric	Cubic smoothing spline based on data from Woodhams *et al.* 2008 Fig. 5.S1E.
Proportion of zoospores contributed to Z by infected adult	Constant	$\mu_A = 1$	Likely a conservative estimate of how infected adults contribute to the zoospore pool
Mean tadpole zoospore load, μ_{T_i}	Constant	$\mu_{T_i} = 1487.036$	Mean zoospore load of tadpoles in mesocosm experiment described in text
Probability of tadpole i surviving a three day time step	Constant	$s_{T_1} = 0.987$; $s_{T_2} = 0.997$; $s_{T_3} = 0.997$	Yearly survival from Briggs *et al.* 2005 converted to a three day time scale
Probability of tadpole i not metamorphosing, p_{T_i}	Constant	$p_{T_1} = 1$; $p_{T_2} = 0.5$; $p_{T_3} = 0$	Briggs *et al.* 2005
Probability of tadpole i surviving metamorphosis, m_{T_i}	Constant	$m_{T_1} = 0.9$; $m_{T_2} = 0.9$; $m_{T_3} = 1.0$	Briggs *et al.* 2005
Probability of adult reproducing, p_A	Constant	$p_A = 0.25$	Briggs *et al.* 2005
Mean adult reproductive output, λ	Constant	$\lambda(x) = \lambda_0 = \lambda = 50$	Leads to realistic population-level growth rate of $\lambda_{\text{growth rate}} \approx 1.46$ in the disease-free, density-independent model (Briggs et al., 2005)
Carrying capacity parameter, K	Constant	$K = 5$	Leads to equilibrium adult densities between 10-15 adults per m^3 in disease-free model

Table 5.2: The results of fitting transmission functions of the form $\phi = 1 - \exp(-\Lambda)$ to the *R. muscosa*-Bd mesocosm experiment. I is the total number of infected adults in a mesocosm at the beginning of a time interval, A is the total number of adults in a tank, Z is the number of zoospores in the mesocosm at the beginning of the time interval as estimated from a latent zoospore pool model (Section 5.A), and Δt is the time between swabbing events in the experiment (between 4-8 days). All β parameters had a half-Cauchy prior from 0 to ∞ with scale parameter equal to 1. Models with lower WAICs and higher weights are better models.

Name	**Function**	**Parameters**	**WAIC (weight)**
Constant zoospore pool	$\Lambda = (\beta_0)\Delta t$	$\beta_0 = 8.07\text{x}10^{-2}$ day^{-1}	401.1 (0)
Density-dependent w/ constant zoospore pool	$\Lambda = (\beta_0 + \beta_1 I)\Delta t$	$\beta_0 = 4.18\text{x}10^{-2}$, $\beta_1 = 5.25\text{x}10^{-2}$	336.7 (0)
Frequency-dependent w/ constant zoospore pool	$\Lambda = (\beta_0 + \beta_1 \frac{I}{A})\Delta t$	$\beta_0 = 4.28\text{x}10^{-2}$, $\beta_1 = 0.551$	336.6 (0)
Dynamic zoospore pool	$\Lambda = (\beta_0 \ln(Z+1))\Delta t$	$\beta_0 = 1.09\text{x}10^{-2}$	345.32 (0)
Density-dependent w/ dynamic zoospore pool	$\Lambda = (\beta_0 \ln(Z+1) + \beta_1 I)\Delta t$	$\beta_0 = 5.29\text{x}10^{-3}$, $\beta_1 = 7.52\text{x}10^{-2}$	**301.18** (0.99)
Frequency-dependent w/ dynamic zoospore pool	$\Lambda = (\beta_0 \ln(Z+1) + \beta_1 \frac{I}{A})\Delta t$	$\beta_0 = 5.77\text{x}10^{-3}$, $\beta_1 = 0.627$	311.06 (0.01)

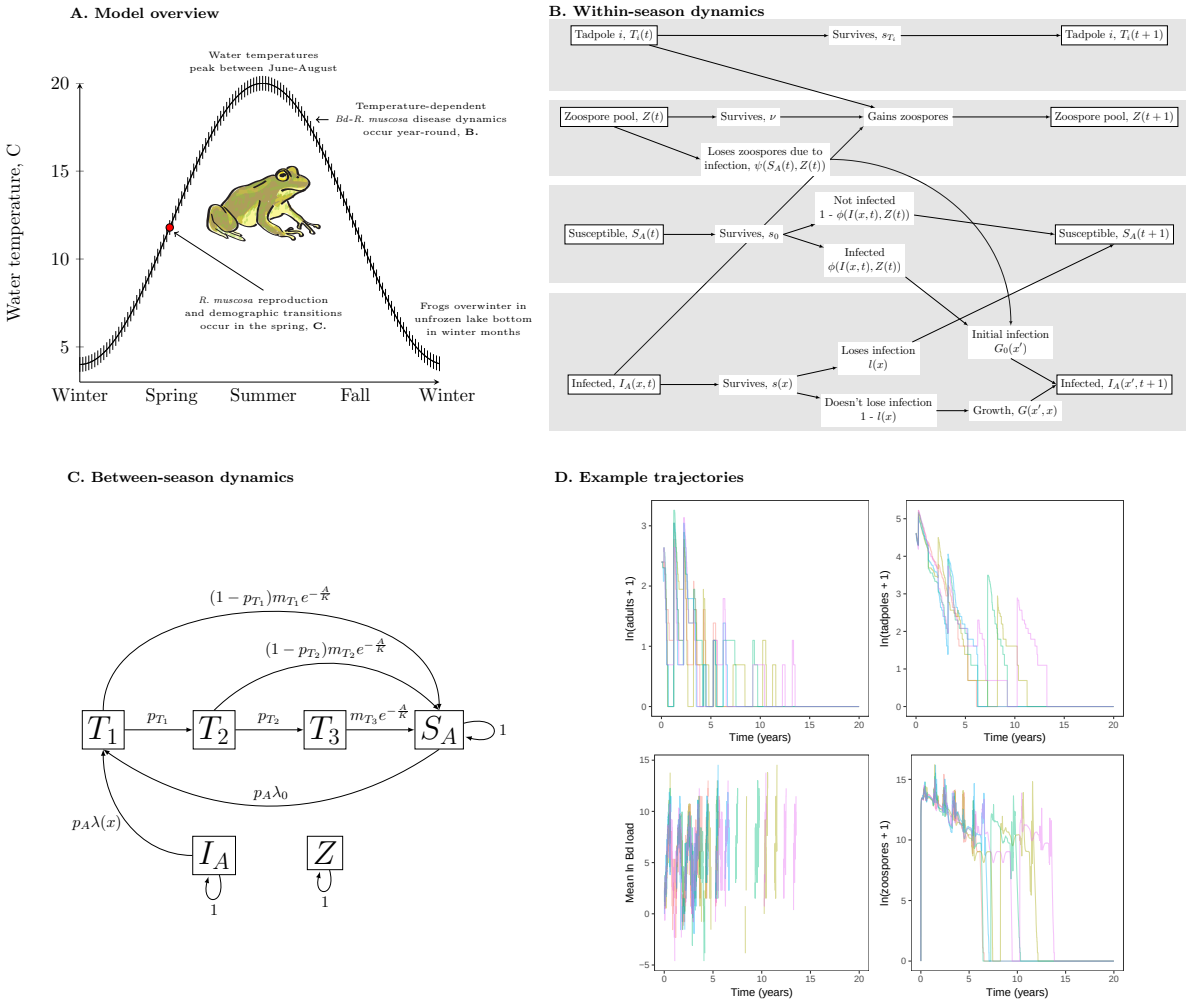


Figure 5.1: **A.** Diagram showing the temporal dynamics of the hybrid model of *R. muscosa*-Bd. Reproduction and demographic transitions occur once a year in the spring (red dot and C.). Disease dynamics are temperature-dependent and are updated every 3 days (vertical dashes and B.) over the course of the entire year. **B.** The within-season dynamics of *R. muscosa* and Bd. **C.** The between season demography of *R. muscosa*. Note that survival probability of susceptible adults (S_A), infected adults (I_A) and Bd zoospores in the environmental pool (Z) is one because their survival probabilities are already accounted for in the within-season model. **D.** Representative trajectories of adult, tadpole and zoospore pool population sizes from the hybrid model with a density-dependent transmission function and infection from a dynamic zoospore pool. Five stochastic trajectories are shown from the hybrid model (colored lines). The mean ln Bd load, conditional on infection, is also shown. Gaps in trajectories for the mean ln Bd load indicate that no infected individuals were in the population at those time points.

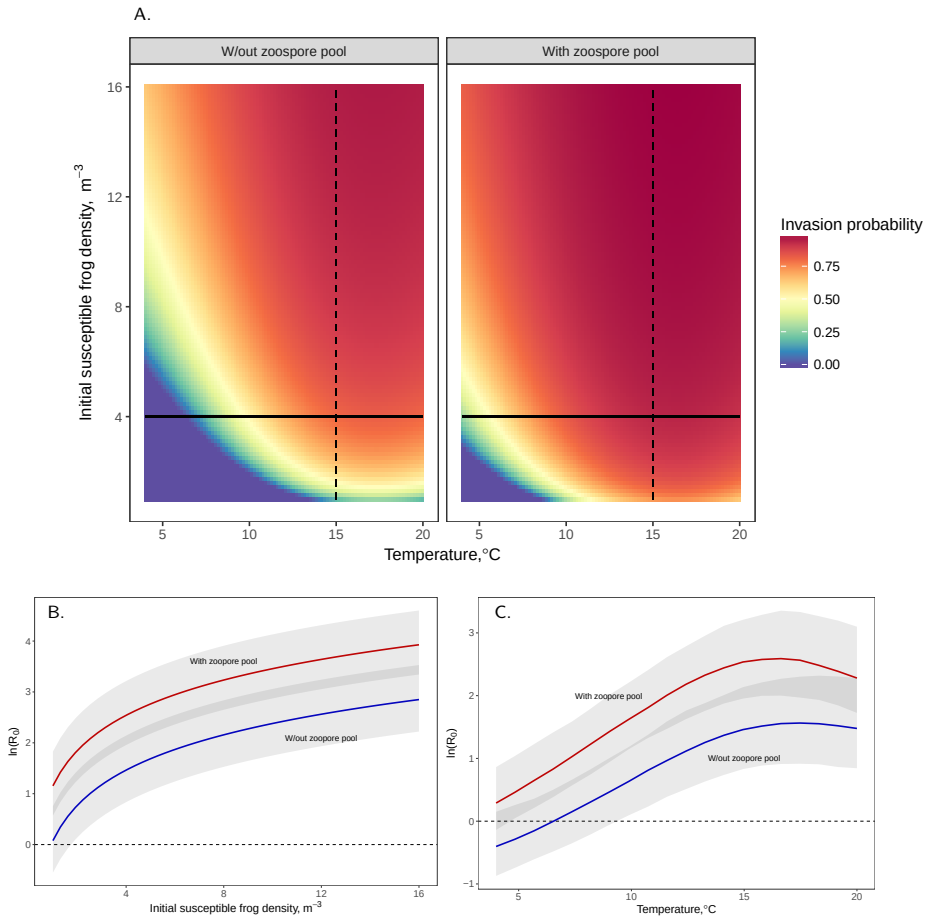


Figure 5.2: R_0 and the invasion probability of Bd $(1 - \frac{1}{R_0})$ for different temperatures and host densities with and without an environmental zoospore pool. This calculation of R_0 uses the parameters given in Table 1 and the "Density dependent w/ dynamic zoospore pool" transmission function in Table 2. While R_0 will inevitably decrease without transmission from a zoospore pool (i.e. when $\beta_0 = 0$; see Section 5.B), the magnitude of that decrease depends on the estimated transmission coefficient from the zoospore pool. **A.** Gives the invasion probability of Bd. The dashed, vertical lines in A. correspond to the curves shown in **B.**, where $\ln(R_0)$ is plotted against initial adult density when temperature is 15 °C. The solid, horizontal lines in A. correspond to the curves shown in **C.** where $\ln(R_0)$ is plotted against temperature when initial adult density is four adults per m^3. The gray regions give the 95% credible intervals. The dashed lines in B. and C. correspond to $R_0 = 1$ $(\ln(R_0) = 0)$, below which Bd cannot invade.

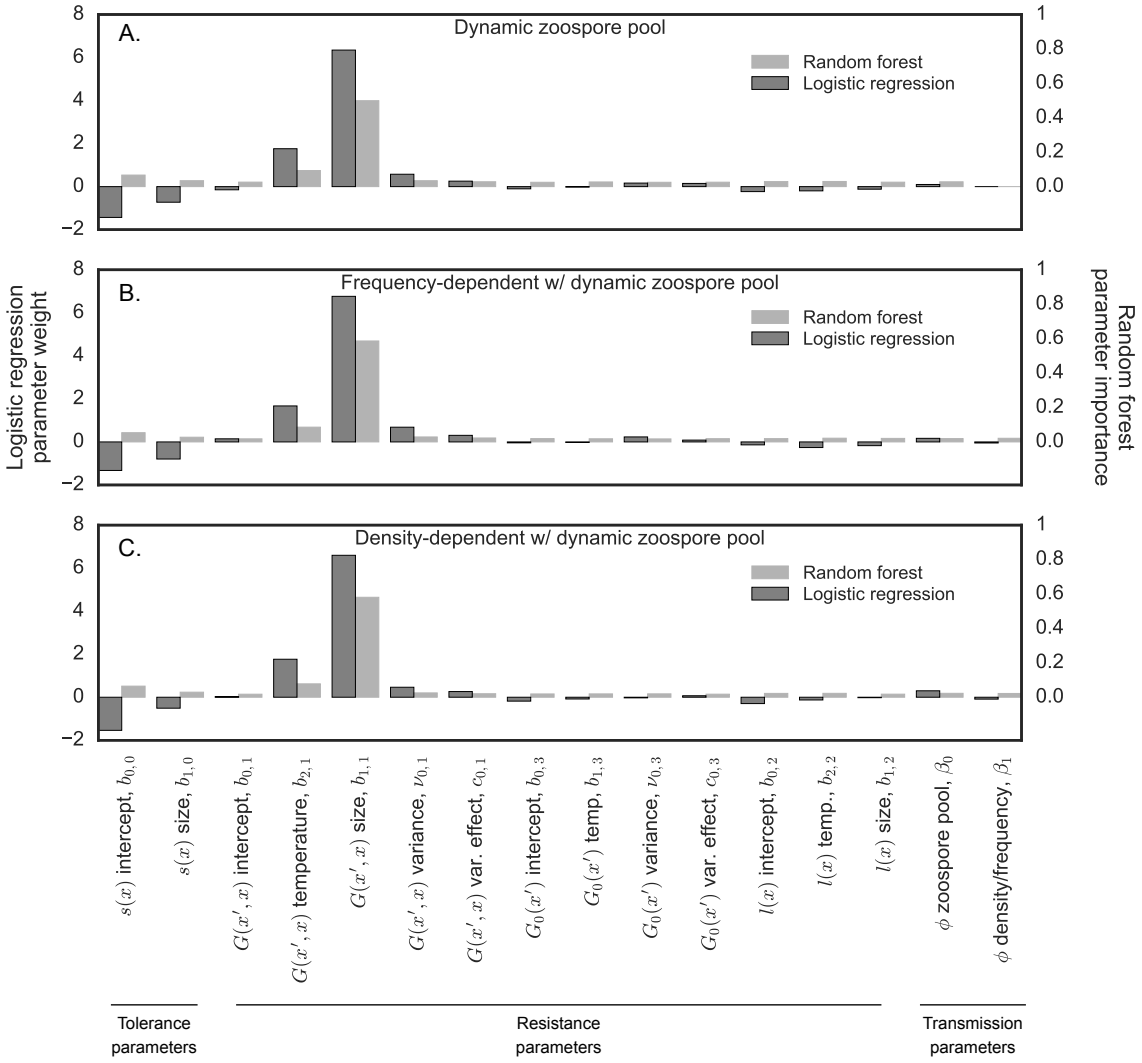


Figure 5.3: **A. - C.** The sensitivity of *R. muscosa* extinction probability to parameters dictating transmission, resistance, and tolerance of Bd for the three transmission functions used in the hybrid model. The dark gray bars give the weights of the various parameters when logistic regression is used to classify whether or not a simulation trajectory experienced extinction. The absolute height of the bar shows the relative importance of that parameter and the direction specifies what happens to the extinction probability when that parameter is increased. For example, increasing the $G(x', x)$ parameter $b_{1,1}$ increases the probability of extinction. The light gray bars give the relative importances of each parameter in predicting the extinction of a simulation when a Random Forest classifier was used to account the interactions between parameters on extinction probability. These values are all between zero and one and the height of a bar indicates the relative importance of a transmission, resistance, or tolerance parameter.

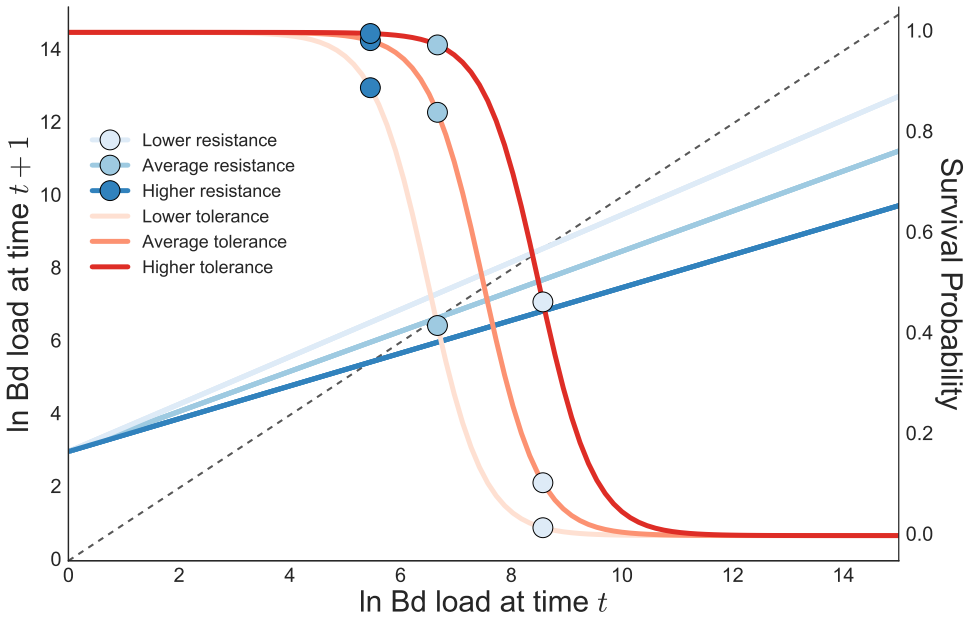


Figure 5.4: The interaction between the growth function determining host resistance (blue lines) and the survival function determining host tolerance (red lines) has large impacts on host survival probability. The solid blue lines of different shades give three different growth functions that describe how Bd load changes from time step t to time step $t+1$. The growth functions have different slopes representing different levels of resistance. When one of the growth functions crosses the dashed 1:1 line, this indicates that the Bd load at time t will be the same as the Bd load at time $t+1$ (i.e. stasis). The red lines of different colors specify different survival functions with different levels of tolerance. The corresponding blue colored circle on one of the red survival curves gives the probability of a host surviving a time step with the given Bd load at stasis. Each red survival curve has three different blue dots for each of the three growth functions. Because of the strong non-linearity of the survival function, changes in the Bd load at stasis can lead to disproportionate changes in survival probability. However, the direction and magnitude of these changes depends on the degree of tolerance represented in the survival curve.

5.A The mesocosm experiment and the latent zoospore pool transmission model

5.A.1 The mesocosm experiment

We performed an outdoor mesocosm experiment in summer 2008 to determine if adult frog density influenced Bd infection prevalence and the rate of increase in the fungal load on individuals through time in Sierra Nevada yellow legged frogs, *Rana sierrae*. The experiment was performed in four 1.2 m wide x 4.8 m long x 1.2 m deep concrete channels at the Sierra Nevada Aquatic Research Laboratory (University of California Natural Reserve System, Mammoth Lakes, CA, N37.614176 W118.833452, elevation 2167m). Channels were subdivided with plywood into 4 isolated tanks measuring 1.2 m^3 in each channel, for a total of 16 tanks in four blocks. Tanks were each filled with approximately 1400 liters of water from adjacent Convict Creek, a Bd-negative stream originating from nearby Convict Lake. Sloping shelves (approximately 30cm wide) extending from below water to a few cm above the water surface were provided on the south-facing wall of each tank for the frogs to bask on during the daylight hours. Lids were constructed from wood frames covered in fiberglass screen and wire mesh.

Infected *Rana sierrae* tadpoles were collected from Conness Pond, a large, Bd-positive site in Yosemite National Park (N37.97485 W119.31134, elevation 3193m). The tadpoles were used as the initial source of infection in each tank. Uninfected (Bd-naïve) adults were collected from Marmot Lake, (Sierra National Forest, N37.259860 W118.683379, elevation 3589m). All animals were swabbed to confirm initial Bd infection status (infected for the tadpoles, and uninfected for the adults). Adult frogs were individually marked with Passive Integrative Transponder (PIT) tags, which allowed each adult to be uniquely identified. On July 7, 2008, five infected tadpoles were added to each tank. Nine days later, tanks were drained 90% and re-filled with fresh water from Convict Creek to remove any tannins leached from the plywood and to remove extra tadpole food. On July 20, 2008, adult frogs were added to each tank in one of four densities: 1, 4, 8, and 16 (this is counted as Day 0 of the experiment for all of the analyses in the main text). The four tanks in each block received one of each of the four density treatments, assigned randomly within each block. Preliminary analyses did not detect a significant effect of block and thus the statistical models presented below do not include a block effect.

Skin swabs were collected from all animals once per week to determine Bd infection status, using

a standardized swabbing protocol (Briggs et al., 2010; Vredenburg et al., 2010). For post-metamorphic individuals, a sterile synthetic swab was brushed across the hind feet (concentrating on the toe webbing), hind legs, and each side of the abdomen, 5 times each, for a total of 30 strokes. For tadpoles, the swab was brushed across the mouthparts for 30 strokes. Swabs were allowed to air dry, and then were placed in individual tubes, and frozen as soon as possible. The number of Bd DNA copies on each swab was determined using quantitative real-time PCR, following the protocol of Boyle et al. (2004), as in Briggs et al. (2010). During each weekly swabbing event, the weight and snout-to-vent length were recorded for each adult. Tadpoles were fed rabbit chow *ad libitum* and adults were fed Phoenix worms *ad libitum* during the swabbing events.

5.A.2 The latent zoospore pool transmission model

Using the mesocosm transmission experiment described above and in the main text, we sought to characterize the transmission function for Bd and *R. muscosa*. In particular, we wanted to quantify how the environmental zoospore pool affected the probability of an amphibian transitioning from uninfected to infected in a time step. However, we were unable to measure the total number of zoospores in the environmental zoospore pool at each time step and therefore assumed that the zoospore pool was a latent variable that needed to be estimated. We estimated it by assuming that the zoospore pool obeyed the following dynamics

$$Z_j(t+\Delta t) = Z_j(t)\exp(-d^*\Delta t) + Z_{j,\text{tadpoles}}(t)f_T^*\Delta t + Z_{j,\text{adults}}(t)f_A^*\Delta t \tag{5.4}$$

This equation assumes that the dynamics of the unobserved zoospore pool in mesocosm j are governed by the survival probability of zoospores in the pool at time t (first term, d^* is zoospore death rate) and contribution of zoospores into the pool based on the total number of zoospores on tadpoles ($Z_{j,\text{tadpoles}}(t)$) and adults ($Z_{j,\text{adults}}(t)$) at time t in mesocosm j. f_T^* and f_A^* are rates that relate the total number of zoospores on tadpoles and frogs at time t (as estimated from qPCR) to the total number of zoospores that enter the pool. This equation does not include the reduction of zoospores in the pool due to frogs acquiring them upon initial infection as we found that the small number acquired is inconsequential for the dynamics of the zoospore pool.

For the ith individual frog in mesocosm j at time $t+1$, the likelihood of gaining an infection is

given by

$$y_{ij}(t+\Delta t) \sim \text{Bernoulli}(\phi_j(t)) \text{ if } y_{ij}(t) = 0 \tag{5.5}$$

Each individual frog i in tank j was swabbed six times over the course of 32 days and $y_{ij}(t)$ specifies whether frog i in mesocosm j was infected (1) or uninfected (0) at time $t+\Delta t$, having been uninfected at time t. Only observations where an animal was uninfected in the previous time step directly contributed to the likelihood ($n = 333$), but the entire time series for all individuals within a mesocosm contributed to the dynamics of the zoospore pool.

$\phi_j(t)$ is the probability of infection in mesocosm j at time t and takes the general functional form

$$\phi_j(t) = 1 - \exp(-\Lambda(Z_j(t), A_j(t), I_j(t))\Delta t) \tag{5.6}$$

where the probability of infection depends on the total number of zoospores in the pool ($Z_j(t)$), the total number of adults ($A_j(t)$), and the total number of infecteds ($I_j(t)$) in tank j at time t. The number of adults and total number of infecteds are both observed (i.e. measured in the experiment) at each time t. In contrast, $Z_j(t)$ is unobserved at each time t. The specific forms of $\Lambda(Z_j(t), A_j(t), I_j(t))$ that we considered are given in Table 2 in the main text. We assumed that transmission was temperature-independent because we were not able to simultaneously manipulate host density and temperature in the mesocosm experiment.

We assumed that the total number of zoospores $Z_j(t)$ is a random variable with a lognormal distribution such that

$$\ln(Z_j(t+\Delta t)) \sim \text{Normal}(\ln(\mu_j(t+\Delta t)) - \frac{\sigma^2}{2}), \sigma^2) \tag{5.7}$$

$$\mu_j(t+\Delta t) = Z(t)_j \exp(-d^*\Delta t) + Z_{\text{tadpoles}}(t)_j f_T^* \Delta t + Z_{\text{adults}}(t)_j f_A^* \Delta t \tag{5.8}$$

$Z_{\text{tadpoles}}(t)_j$ and $Z_{\text{adults}}(t)_j$ were both observed at each time t. The $\frac{\sigma^2}{2}$ terms is a result of converting from the expected value of the lognormal distribution to the mean of the normal distribution on the log scale.

To fit these models, we assumed that the initial zoospore load in the pool was given by $\ln Z_j(t = 0) \sim \text{Normal}(\ln(2000), \sigma^2)$. We specified this prior distribution based on the average number zoospores

on the tadpoles at the beginning of the experiment across all mesocosms. We gave the zoospore death rate d^* day^{-1} a normal prior distribution with mean 0.3 day^{-1}, standard deviation 0.03, and a lower bound of 0. This tight prior was based on laboratory estimates of the zoospore death rate (Woodhams et al., 2008). To aid in the identifiability of our model, we set $\sigma = 1$ and $f_T = f_A = f$ and gave f a vague half Cauchy prior distribution with a scale parameter equal to 1.

We fit this model for each of the ϕ functions given in Table 2 in the main text using Hamiltonian Monte Carlo with the RStan package (2.12.1). Three chains were run for each model and we assessed convergence of the model parameters by determining if the Gelman-Rubin statistic $\hat{R}$ was less than 1.05 (Gelman et al., 2014). We also confirmed that this statistical model could recover known transmission functions by simulating our mesocosm experiment *in silico* and testing whether the above model could both recover known parameters in ϕ and also correctly distinguish between different forms of ϕ using information criteria (see accompanying code `code/simulate_lab_data.py` at `https://github.com/mqwilber/host_extinction/`).

We also tested how robust our conclusions were to the assumption that $\sigma = 1$ using two different approaches. First, we allowed this parameter to have a minimally informative half Cauchy prior with a scale parameter equal to three. This allowed for enormous variability in the dynamics of the zoospore pool. Incorporating this vague prior led to larger mean estimates of transmission from the zoospore pool (i.e. larger β_0), but also larger uncertainty around this estimate. However, the relative ranking of the different transmission models given in Table 2 in the main text did not change. The models with a dynamic zoospore pool were always better than corresponding models without the zoospore pool and the density-dependent model with a dynamic zoospore model was the best model followed by the frequency-dependent model with a zoospore pool.

Second, we explored how fixed values of σ ranging from 0.25 to 4 affected the relative rank of the models as well as the estimates of the coefficients (Figure 5.S9, 5.S10). Across a range of values of σ and two different measures of information criteria (WAIC and DIC), the relative rankings of the three transmission models with a dynamic zoospore pool (Table 2 in main text) stayed largely consistent, with some notable deviations between $\sigma = 1.5$ and $\sigma = 2$ (Fig. 5.S9). We used two different information criteria as there is some question over WAIC's accuracy for time series models (Gelman et al., 2014). For values of $\sigma \leq 1$ the estimates of the transmission coefficients from the zoospore pool (β_0) remained relatively constant, but began increasing for $\sigma > 1$ due to the large variability in the zoospore pool

dynamics (Fig. 5.S10A). In contrast, the coefficient estimates for the density/frequency dependent transmission stayed relatively constant with a slight decreasing trend with increasing σ (Fig. 5.S10B). Considering these sensitivity analyses, we felt that $\sigma = 1$ was a reasonable choice because 1) it did not lead to unrealistically large variability in the zoospore pool as did the uninformative prior on σ 2) it provided a conservative estimate for effect of the zoospore pool on transmission 3) the density/frequency dependent transmission coefficient β_1 was largely unaffected by the choice and 4) it resulted in model ranks that were consistent between two different information criteria and largely consistent across most values of σ that we explored.

5.B R_0 for host-parasite IPMs with an environmental reservoir

5.B.1 R_0 for a discrete-time SIS model with an environmental reservoir

Wilber et al. (2016) showed how to calculate R_0 for basic host-parasite IPMs using the general methodology developed in Klepac and Caswell (2011). This method can be extended to calculate R_0 for an IPM with an environmental reservoir. To begin, we illustrate the procedure with a simple discrete time Susceptible-Infected-Susceptible (SIS) model with a dynamic zoospore pool (Z). Take the following set of discrete dynamical equations

$$S(t+1) = S(t)s_0[1 - \phi(I(t), Z(t))] + I(t)s_I l_I \tag{5.9}$$

$$I(t+1) = I(t)s_I(1 - l_I) + S(t)s_0\phi(I(t), Z(t)) \tag{5.10}$$

$$Z(t+1) = Z(t)\nu + I(t)f \tag{5.11}$$

where s_0 is the survival probability for an uninfected host in a time step, s_I is the survival probability for an infected host in a time step, l_I is the probability of losing an infection in a time step, ν is the survival probability of a zoospore in a time step, f is the average number of zoospores produced by an infected individual in a time step, and $\phi(I(t), Z(t))$ is the probability of gaining an infection in a

time step. Based on the results from the transmission experiment given in Table 2 in the main text, we assume that transmission depends on density-dependent host to host contact as well as transmission from the zoospore pool. Given this, we can write

$$\phi(I(t), Z(t)) = 1 - \exp[-(\beta_1 I(t) + \beta_0 Z(t))] \tag{5.12}$$

We can calculate R_0 for the above system of equations by rewriting them as the following matrix model

$$\begin{bmatrix} S \\ \mathbf{P} \end{bmatrix}(t+1) = \begin{bmatrix} 0 & \mathbf{0} \\ \mathbf{M}(\mathbf{P}(t)) & \mathbf{U} \end{bmatrix} \begin{bmatrix} S \\ \mathbf{P} \end{bmatrix}(t) \tag{5.13}$$

where the zeros simply indicate that these values do not contribute to the calculation of R_0, not that they are actually zero in the model. $\mathbf{P}(t)$ is the vector $[I(t)\ Z(t)]^T$ and $\mathbf{0}$ is a row vector [0 0]. Just focusing on the vector $\mathbf{P}$, we can write

$$\mathbf{P}(t+1) = \mathbf{M}(\mathbf{P}(t))S(t) + \mathbf{U}\mathbf{P}(t) \tag{5.14}$$

$\mathbf{M}(\mathbf{P}(t))$ is the vector $[s_0\phi(\mathbf{P}(t))\ 0]^T$. $\mathbf{U}$ is the 2 x 2 matrix

$$\mathbf{U} = \begin{bmatrix} s_I(1-l_I) & 0 \\ f & \nu \end{bmatrix} \tag{5.15}$$

To calculate R_0, we linearize $\mathbf{P}(t+1)$ about a vector $\mathbf{n}^*$. We set this vector to be a host population with only susceptibles $\mathbf{n}^* = [S^*\ \mathbf{0}]$ (Rohani et al., 2009; Klepac and Caswell, 2011), where $\mathbf{0}$ is a vector of zeros of length $n = 2$. We then compute the Jacobian matrix evaluated at $\mathbf{n}^*$

$$\mathbf{J} = \left.\frac{d\mathbf{P}(t+1)}{d\mathbf{P}(t)}\right|_{\mathbf{n}^*} \tag{5.16}$$

which allows us to compute R_0 (Klepac and Caswell, 2011). Computing this Jacobian requires computing $\dfrac{d\mathbf{M}(\mathbf{P}(t))S(t)}{d\mathbf{P}(t)}$ and $\dfrac{d\mathbf{U}\mathbf{P}(t)}{d\mathbf{P}(t)}$.

We immediately see that $\left.\dfrac{d\mathbf{U}\mathbf{P}(t)}{d\mathbf{P}(t)}\right|_{\mathbf{n}^*} = \mathbf{U}$. $\left.\dfrac{d\mathbf{M}(\mathbf{P}(t))S(t)}{d\mathbf{P}(t)}\right|_{\mathbf{n}^*}$ requires application of the chain rule

for differentiation and we see that

$$\left.\frac{d\mathbf{M}(\mathbf{P}(t))S(t)}{d\mathbf{P}(t)}\right|_{\mathbf{n}^*} = \mathbf{M}^* = \begin{bmatrix} s_0 S^* \beta_1 & s_0 S^* \beta_0 \\ 0 & 0 \end{bmatrix} \tag{5.17}$$

R_0 is given by (Klepac and Caswell, 2011)

$$R_0 = \max \text{ eig}(\mathbf{M}^*(\mathbb{1} - \mathbf{U})^{-1}) \tag{5.18}$$

and plugging into $\mathbf{M}^*$ and $\mathbf{U}$ given above we get

$$R_0 = \frac{s_0 S^* \beta_1}{1 - s_I(1 - l_I)} + \frac{s_0 S^* \beta_0 f}{(1-\nu)[1 - s_I(1 - l_I)]} \tag{5.19}$$

Similar to the continuous time result given in Rohani et al. (2009), we see that R_0 is a combination of both transmission from the hosts (first term) and the environment (second term). Setting $\beta_0 = 0$, we recover the simple SIS R_0 given in Wilber et al. (2016) and Oli et al. (2006).

5.B.2 R_0 for the host-parasite IPM with an environmental reservoir

To generalize this to a host-parasite IPM model with an environmental reservoir where the I class is now potentially infinitely many classes, we can take the following steps. First, we recognize that in practice IPMs are analyzed using the mid-point rule (Easterling et al., 2000) such that there are a finite number of I classes, namely n classes. Therefore, we can think of our IPM as a generalization of the SISZ model presented above such that $\mathbf{P}(t)$ is a vector of length $n + 1$, $\mathbf{M}(\mathbf{P}(t))$ is now a vector of length $(n+1)$ and $\mathbf{U}$ is a $(n+1)$ x $(n+1)$ matrix. The plus one is because the environmental reservoir Z is part of the $\mathbf{P}$ vector.

For the first n x n elements in $\mathbf{U}$, the element in the ith row and the jth column is given by

$$u_{ij} = s(x_j)(1 - l(x_j))G(x_i, x_j)\Delta \tag{5.20}$$

which gives the probability of an individual in the jth load class with parasite load x_j, surviving $(s(x_j))$, not losing its infection $(1 - l(x_j))$, and transitioning to the load class of x_i in a time step $(G(x_i, x_j))$.

Δ is a result of using the midpoint rule to discretize the continuous IPM and is used to convert the probability density $G(x_i, x_j)$ into a probability (e.g. see Easterling et al., 2000; Wilber et al., 2016). The $n+1$th row of $\mathbf{U}$ is the vector $[f_x\mathbf{x} \;\; \nu]$ of length $n+1$, where we assume that an infected host in class j produces an average of $f_x x_j$ parasites in a time step into the zoospore pool. $\mathbf{x}$ is a vector of length n that contains the corresponding parasite loads for the n infected classes. The $n+1$th column of $\mathbf{U}$ is the vector $[\mathbf{0} \;\; \nu]^T$, where $\mathbf{0}$ is of length n.

$\mathbf{M}(\mathbf{P}(t))$ is given by a vector of length $n+1$ where the first $i = 1, \ldots, i, \ldots n$ elements are given by

$$m_i = s_0\phi(\boldsymbol{\beta}^T\mathbf{I}(t))G_0(x_i)\Delta \tag{5.21}$$

where $\boldsymbol{\beta}^T$ is a vector of length $n+1$ with the first n elements being β_1 and the $n+1$th element being β_0. $\left.\frac{d\mathbf{M}(\mathbf{P}(t))S(t)}{d\mathbf{P}(t)}\right|_{\mathbf{n}^*} = \mathbf{M}^*$ is given by a $(n+1)$ x $(n+1)$ matrix where the first n columns are given by $[s_0S^*\beta_1G_0(\mathbf{x})\Delta \quad 0]^T$ and the $n+1$th column is given by $[s_0S^*\beta_0G_0(\mathbf{x})\Delta \quad 0]^T$. $G_0(\mathbf{x})$ indicates that the function $G_0(x)$ is evaluated at each element in $\mathbf{x}$, resulting in a vector of length n.

Given $\mathbf{M}^*$ and $\mathbf{U}$ we can again calculate R_0 as (Klepac and Caswell, 2011)

$$R_0 = \max \text{ eig}(\mathbf{M}^*(\mathbb{1} - \mathbf{U})^{-1}) \tag{5.22}$$

In the main text, we use this formulation to numerically calculate R_0 for the *R. muscosa*-Bd IPM model with and without an environmental reservoir.

5.C The hybrid model

5.C.1 Within-year component of the hybrid model

The model described by equations 6-8 in the main text may be sufficient to describe the dynamics of an initial epizootic, but in order to examine Bd-induced extinction dynamics in *R. muscosa* populations a number of additions need to be made. First, the tadpole stage of *R. muscosa* has been shown to play an important role in generating enzootic dynamics in *R. muscosa* populations (Briggs et al., 2005, 2010). *R. muscosa* can spend three years as tadpoles and thus we include three additional tadpole stages into the model T_1, T_2, and T_3 (Briggs et al., 2005). Considering equations 6-8, we can add this tadpole class and update our zoospore pool equation as follows

$$T_i(t+1) = T_i(t)s_{T_i} \tag{5.23}$$

$$Z(t+1) = Z(t)\nu + \sum_{i=1}^{3} \mu_{T_i} T_i(t) + \mu_A \int_{L_x}^{U_x} \exp(x) I_A(x,t) dx - \psi(S_A(t), Z(t)) \tag{5.24}$$

The equation $T_i(t+1)$ describes how the number of tadpoles in class i changes from time t to time $t+1$. We do not explicitly model the fungal load on tadpoles. Instead, we assume all tadpoles are immediately infected with Bd and have a constant contribution to the zoospore pool. This is justified by the observation that most tadpoles in *R. muscosa* populations carry high fungal loads, even in enzootic populations (Briggs et al., 2010). *R. muscosa* tadpole survival is not affected by Bd infection. Therefore, the within-season dynamics of T_i are simply given by the probability of a tadpole surviving from time t to $t+1$, which is s_{T_i}. Notice that infected tadpoles are now also contributing μ_{T_i} zoospores to the zoospore pool at each time step. Previous models of this system have included a subadult stage after metamorphosis that can last 1-2 years (Briggs et al., 2005, 2010). For simplicity, we are ignoring it here.

R. muscosa populations also experience seasonal temperature fluctuations in which lake temperatures drop to approximately 4 °C in the winter (in the unfrozen portion of a lake where the frogs overwinter) and reach approximately 20 °C in the summer (Knapp et al., 2011). We account for this seasonal variability by imposing a deterministically fluctuating environment on the *R. muscosa*-Bd IPM. We assumed that temperature follows a sinusoidal curve with a period of 1 year and a minimum temperature of 4 °C and a maximum temperature of 20 °C (Fig. 5.1A). At each discrete time point within a season, a new temperature is calculated based on the sinusoidal curve and the temperature-dependent vital rate functions are updated accordingly (see Table 1 in the main text and Fig. 5.S1 for temperature-dependent vital-rate functions).

5.C.2 Between-year component of the hybrid model

In the within-year component of the hybrid model a time step is 3 days. This time step is on the same scale as Bd dynamics. However, *R. muscosa* demography occurs on a slower scale. We assume that *R. muscosa* demographic dynamics occur on a yearly time scale (Briggs et al., 2005), such that once a year tadpoles either age a year or metamorphose and adults reproduce. The mortality for each

stage as well as all disease dynamics are accounted for in the within-year component of the IPM, thus the between-year component only includes the following demographic events (Fig. 5.1C in the main text)

$$T_1(t+1) = p_A \lambda_0 S_A(t) + p_A \int_{L_x}^{U_x} \lambda(x) I(x,t) dx \tag{5.25}$$

$$T_2(t+1) = p_{T_1} T_1(t) \tag{5.26}$$

$$T_3(t+1) = p_{T_2} T_2(t) \tag{5.27}$$

$$S_A(t+1) = [(1-p_{T_1}) m_{T_1} T_1(t) + (1-p_{T_2}) m_{T_2} T_2(t) + m_{T_3} T_3(t)] \exp(-A(t)/K) \tag{5.28}$$

$$I_A(x', t+1) = I_A(x', t) \tag{5.29}$$

$$Z(t+1) = Z(t) \tag{5.30}$$

where these events only occur once per year (Fig. 5.1 in the main text). When these demographic events occur each year, they occur after the disease dynamics defined above, but in the same time step from t to $t+1$. Equation 22 gives the contribution of adult frogs to the T_1 tadpole class. p_A defines the probability of an adult reproducing, λ_0 is the mean reproductive output of uninfected adults and $\lambda(x)$ is the mean reproductive output of an adult with a Bd load of x. Equation 23 gives the probability of a T_1 tadpole not metamorphosing (p_{T_1}) and transitioning to a T_2 tadpole. Equation 24 describes the changes in the T_3 class via the probability of a T_2 tadpole not metamorphosing (p_{T_2}) and transitioning to a T_3 tadpole. Equation 25 describes T_1, T_2 and T_3 tadpoles metamorphosing, surviving metamorphosis, and recruiting as uninfected adults. Recruitment of tadpoles to the adult stage is a density-dependent process following a Ricker function with adult density ($A(t)$) and a parameter that is proportional to the carrying capacity (K) (Briggs et al., 2005). Because we have no empirical evidence for Bd-induced fertility reduction in *R. muscosa*, we assumed that reproduction in uninfected adults was the same as reproduction in infected adults. We also assume that tadpoles lose their infection during Bd metamorphosis (Briggs et al., 2010). Finally, equations 26 and 27 indicate that infected individuals ($I_A(x', t+1)$) and the zoospore pool ($Z(t+1)$) do not change during the demographic update.

5.D Converting the hybrid model into an individual-based model with demographic stochasticity

5.D.1 Details of the individual-based model

The simplest way to include demographic stochasticity into the hybrid model is to assume that demographic stochasticity is given by sampling error and allow all demographic transitions to occur following some probability distribution (Caswell, 2001; Schreiber and Ross, 2016). To illustrate this, recall that to analyze the hybrid Integral Projection Model we discretized the continuous class $\int_a^b I(x,t)dx$ (which gives the number of frogs with a ln Bd load between a and b at time t) into 30 load classes using the midpoint rule (Easterling et al., 2000). We can now loosely think about our hybrid model as a matrix model with 3 tadpole + 1 susceptible adult + 30 infected adult + 1 zoospore pool = 35 stages. Let's call the transition matrix $\mathbf{A}$. Following Caswell (2001) and Schreiber and Ross (2016), we can decompose our hybrid model into components for infection dynamics (i.e. growth, loss, initial infection gain), host survival, disease transmission, and demographic transitions ($\mathbf{T}$) and reproduction ($\mathbf{F}$) such that $\mathbf{A} = \mathbf{T} + \mathbf{F}$. The matrix $\mathbf{T}$ gives the probabilities of an individual in stage j transitioning to stage i in a 3 day time step. However, $\mathbf{T}$ is not a stochastic matrix (i.e. the columns do not sum to one) because individuals in each stage also have a survival probability and one potential transition during each time step is to a "dead" class. We can augment this matrix $\mathbf{T}$ with an extra row $\mathbf{d}$ specifying the "dead" class where each entry in this row can be defined as $d_{36,j} = 1 - \text{sum}(\mathbf{T}_{,j})$. $\text{sum}(\mathbf{T}_{,j})$ gives the sum of the jth column of the $\mathbf{T}$ matrix. The new augmented matrix $\mathbf{T}^*$ is fully stochastic (Caswell, 2001; Schreiber and Ross, 2016). Assuming each individual transitions independently of each other, for each time step t the transition of a single individual in class j to another class i (including the "dead" class) follows a multinomial distribution with a probability vector $\mathbf{p_j}$ given by the jth column of $\mathbf{T}^*$: $\mathbf{p_j} = \mathbf{T}^*_{,\mathbf{j}}$. At any time step we could simulate what happens to n individuals in class j by drawing from a multinomial distribution with the total number of trials equal to n and the probability vector equal to $\mathbf{p_j}$. The resulting random vector would be of length 36 specifying the stages that the n individuals of stage j at time t now occupy at time $t+1$ (including death).

In addition to individual transitions (or "births"), we also need to account for the births defined in the $\mathbf{F}$ matrix. In the hybrid model, there are two types of births we account for in the $\mathbf{F}$ matrix: the

production of T_1 tadpoles from adults (infected and uninfected) and the production of zoospores from tadpoles and infected adults. As is, the fecundity matrix $\mathbf{F}$ only specifies the mean production of T_1 tadpoles and zoospores. To make this probabilistic, we need to specify a distribution around this mean. We assumed that the production of tadpoles followed a Poisson distribution with mean $A(t)p_A\lambda$, where $A(t) = S_A(t)+\int_{L_x}^{U_x} I_A(x,t)dx$ is the number of adults in the population at time t, p_A is the probability of an adult reproducing and λ is the mean number of tadpoles produced. We used a Poisson distribution for reproduction for consistency with previous modeling studies of *R. muscosa* (Briggs et al., 2005, 2010). Moreover, as our goal was to understand the relative effects of resistance, tolerance, and transmission on extinction risk, the exact form of the reproduction distribution has little influence of the effects of these three process relative to each other. In contrast, the shape of the reproductive distribution will have enormous implications if one were to consider the evolution of resistance and tolerance in a host population. While this was beyond the scope of this study, an important next step is to understand how standing variation in resistance and tolerance can rescue populations from extinction (Orr and Unckless, 2014). For this, particular attention will need to be paid to both the reproductive distribution and any trade-offs between increased resistance/tolerance and mean host reproduction (Boots et al., 2009).

We also assumed that the number of zoospores shed into the zoospore pool from tadpoles and adults at each time step followed a Poisson distribution with mean $\sum_{i=1}^{3} \mu_{T_i} T_i(t) + \mu_A \int_{L_x}^{U_x} \exp(x) I_A(x,t)dx$. The Poisson distribution is justified by first noting that the best estimate for the rate of zoospore shedding from an adult or tadpole at any point in a time step is the mean shedding rate over the entire time step. Given no other information, the best guess is to assume a constant rate of zoospore shedding over a time step. Taking this constant rate of zoospore shedding over a time step, probability theory then tells that we should expect a Poisson distribution of zoospores produced per time step (Grimmett and Stirzaker, 2001). If the rate of zoospore shedding actually varies over a time step, we would no longer expect a Poisson distribution of zoospores released per individual over a time step. However, because we have no information on how zoospore shedding rate varies on a per individual basis over a three-day time step, the Poisson distribution provides a reasonable null hypothesis.

With these distributional assumptions, we can then calculate the contribution of T_1 tadpoles and zoospores over a time step t as a draw from a Poisson distribution with a mean given by the appropriate entry in the $\mathbf{F}$ matrix. Combining these draws from a Poisson distributions with the draws from the multinomial distribution described above, we can simulate the individual-based representation of our

hybrid model with any density-dependent assumptions about recruitment or disease transmission in addition to any assumptions about temperature-dependent vital rates. This can be done by updating the various transition probabilities using the appropriate densities or temperatures at each time step and performing the previously described stochastic draws with the updated transition probabilities. See the function `multiseason_simulation` in `IPM_functions_for_R.R` for the R code necessary to implement this stochastic simulation (all code available at `https://github.com/mqwilber/host_extinction`).

5.D.2 Testing the assumptions regarding how tadpoles contribute zoospores to the zoospore pool

In the above stochastic model, we assumed that tadpoles were always infected and shed a Poisson-distributed random variable Z_{shed} of zoospores into the zoospore pool at each time step, where Z_{shed} has a constant mean μ_T. This was a simplifying assumption of our model based on field data showing that tadpoles have high Bd prevalence, even in endemic populations (Briggs et al., 2010). However, this prevalence tends to be less than 1 (e.g. ≈ 0.84 in the mescosm experiment and between 0.3 and 1 at certain sites the field, Briggs et al., 2010), which is not exactly consistent with what we assumed in the model. To test the influence of our assumption that tadpole prevalence is one, it would be ideal to explicitly model tadpole load dynamics as we do with adult load dynamics. However, this was not possible as we could not uniquely identify tadpoles in the mesocosm experiment as they were too small to PIT tag.

Instead, we modified our assumption that an individual tadpole contributed $Z_{\text{shed}} \sim \text{Poisson}(\mu_T)$ zoospores to the zoospore pool per time step. We allowed an individual tadpole to contribute $Z_{\text{shed}} \sim$ Zero-inflated Poisson(μ_T, p) zoospores to the zoospore pool per time step, where p gives probability that a tadpole is infected. The justification for the Poisson distribution is the same as given in the previous section, but now we allow some tadpoles to be uninfected in a time step with probability p, such that they do not contribute to the zoospore pool in that time step. This modification allows us to relax our assumption that all tadpoles are infected in any given time step, without having to explicitly model tadpole infection dynamics. Based on the data from the mesocosm experiment used in this study, we set $p = 0.84$ and $\mu_T = 1487.036$.

Rerunning our model with the above changes, we found that allowing for tadpole prevalence to be

the empirically-derived value of $p = 0.84$ did not qualitatively affect the extinction dynamics in this system (Fig. 5.S5). The reason for this is because infected tadpoles were still able to shed a sufficient number of zoospores to keep the force of infection in the system high. To understand when reduced tadpole Bd prevalence did begin to affect host extinction dynamics, we reduced tadpole prevalence to 0.1, which is far lower than the observed levels of tadpole Bd prevalence in this system (Briggs et al., 2010). This low prevalence did reduce extinction risk relative to the baseline model and the model with prevalence set to 0.84, but had less of an effect on extinction risk than increasing the rate of zoospore decay in the environment (Fig. 5.S5). Given these additional analyses showing that large changes in tadpole prevalence had small effects on the extinction dynamics, we felt confident that our simplifying assumption that tadpoles were always infected did not influence our main conclusions.

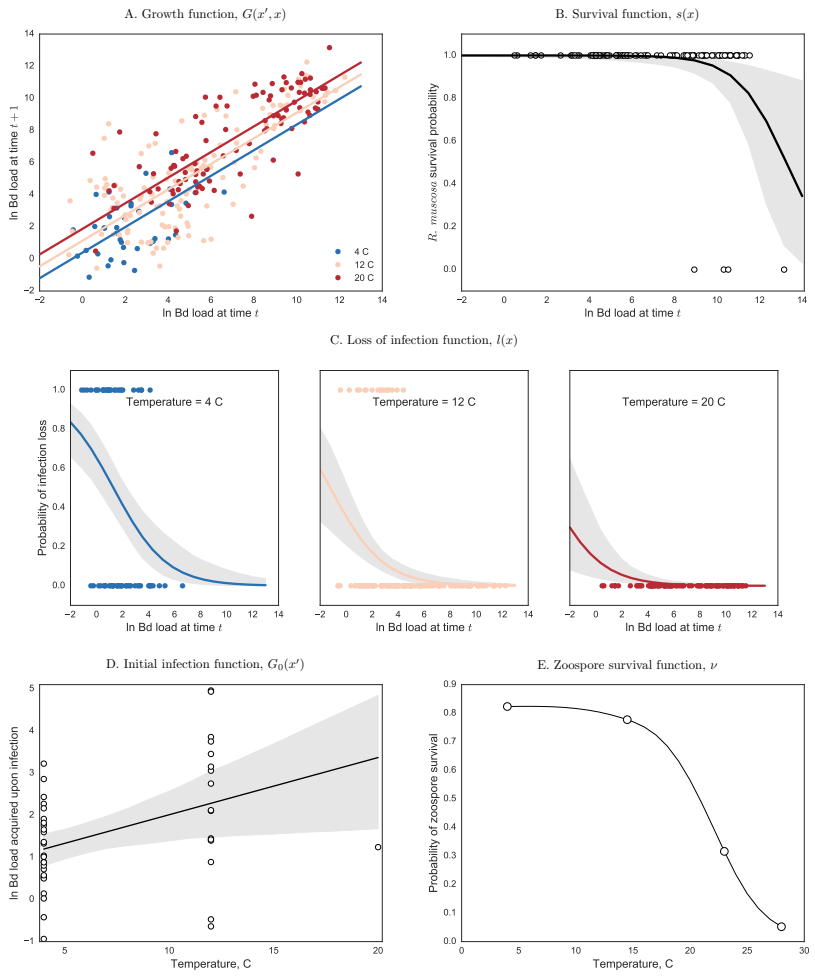


Figure 5.S1: The various vital rate functions estimated from laboratory data given in Wilber et al. (2016) and Woodhams et al. (2008). In all plots, points give the laboratory data, lines give the model fit, and gray regions given the 95% credible interval about the predictions. **A.** The temperature-dependent growth function $G(x', x)$ where temperature is included as a continuous covariate in the growth function. 95% CIs where not included for visual clarity. **B.** The survival function $s(x)$ which dictates the probability of an adult *R. muscosa* surviving with a given load over a three day time step. **C.** The loss of infection function $l(x)$ which gives the load- and temperature-dependent probability of an adult *R. muscosa* losing a *Bd* infection over a three day time step. **D.** The initial infection function $G_0(x')$ that specifies the temperature-dependent probability density of gaining an initial infection of size x'. **E.** The zoospore survival function ν that gives the temperature-dependent probability of a zoospore surviving over three days as given in Woodhams et al. (2008). No uncertainty was included around this prediction.

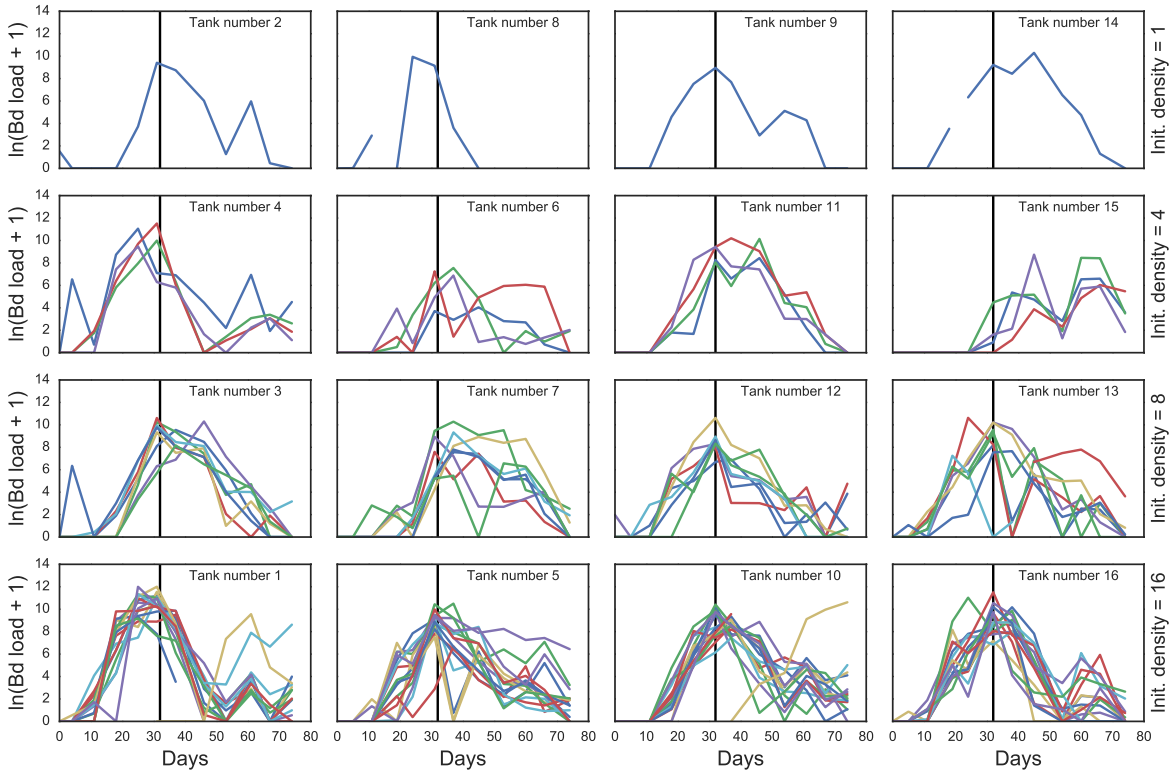
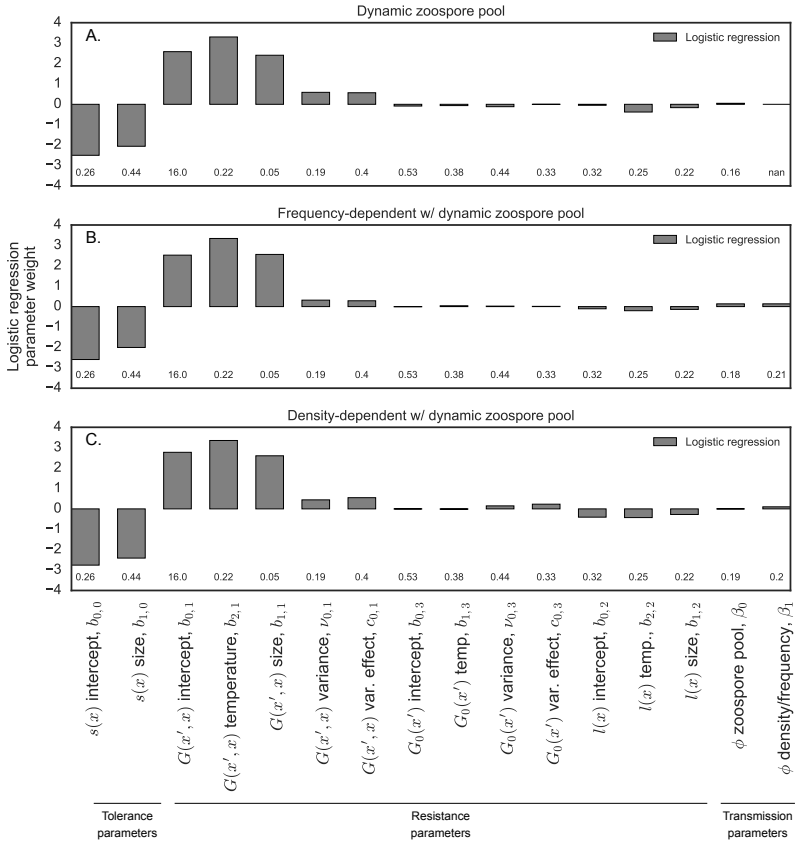


Figure 5.S2: The observed *Bd* load trajectories for individual frogs in the density-dependent mesocosm experiment described in the main text. Tanks/mesocosms had an initial density of either 1, 4, 8, or 16 frogs and were assigned numbers 1 - 16 as indicated on a panel. The different colored lines are the different *Bd* load trajectories for the frogs in given tank. The black vertical line indicates day 32 of the experiment, after which there was an unexplained decrease in zoospore load in the infected frogs in all mesocosms in which frogs were infected by this time point. The consistency of the decline between treatments, between mesocosms, and between frogs suggests the involvement of an external environmental driver or a frog immune response. Data after this time point was not used when fitting the transmission models described in the main text.



;

Figure 5.S3: The results of the global sensitivity analysis similar to Figure 5.4 in the main text. However, instead of perturbing parameters by drawing from a lognormal distribution with median 1 and $\sigma = 0.3$, we drew the parameters from their corresponding posterior distribution from the fitted Bayesian models of the temperature and transmission experiments described in the main text. This allowed for correlation between the parameters as well as allowing some parameters to have larger variability than others, strictly based on the variability in the posterior distribution. The above plot gives the standardized coefficients (i.e. all predictors were z-transformed before the analysis) from a regularized logistic regression. The height of a bar indicates how sensitive extinction risk was to this parameter. The direction indicates whether increasing the parameter increased or decreased the probability of extinction. The values under each bar give the coefficient of variance (CV) for each parameter, calculated from that parameter's posterior distribution ("nan" indicates this parameter was not in the model). The results from this sensitivity analysis were qualitatively consistent with the approach used in the main text: disease-induced extinction was far more sensitive to the parameters in the growth function ($G(x', x)$, resistance) and the survival function ($s(x)$, tolerance) than to parameters in the transmission function. Pruned regression trees also confirmed the important interaction between the growth function and the survival function.

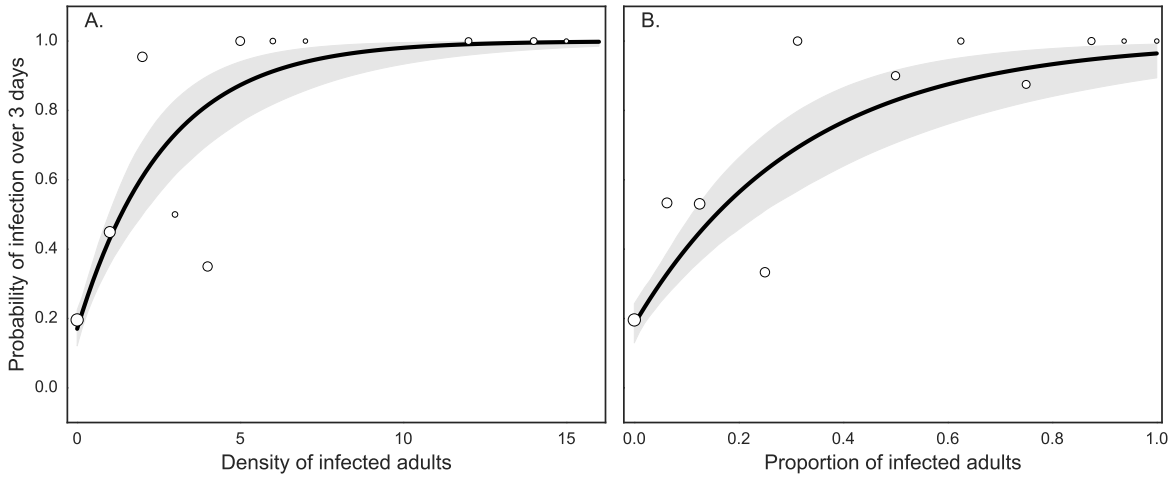


Figure 5.S4: Predictions from the best fit frequency-dependent and density-dependent transmission functions with a dynamic zoospore pool, as estimated from the mesocosm experiment described in the main text. Because the data used to fit these transmission models are Bernoulli (0 or 1) and there are three dimensions of predictor variables (time, host density/frequency, and zoospore density) it is difficult to give a visual representation of model fit. Here we show the marginal predictions (black lines) from the frequency-dependent and density-dependent transmission models, fixing the time step at 3 days and the zoospore pool at 1096 zoospores. The gray region gives the 95% credible interval around these predictions. The open circles give the observed proportion of frogs that transitioned from uninfected to infected with the proportion/density of infected individuals given on the x-axis (pooled across different time steps and zoospore pool sizes). The size of the points indicates the sample size of each point, with larger points indicating larger sample size ($n = 333$ total samples distributed among the points). **A.** The predictions from the best-fit density-dependent transmission function ($\phi = 1 - \exp\left(-(\beta_0 \ln(Z+1) + \beta_1 I)\Delta t\right)$) **B.** The predictions from the best-fit frequency-dependent transmission function ($\phi = 1 - \exp\left(-(\beta_0 \ln(Z+1) + \beta_1 \frac{I}{A})\Delta t\right)$)

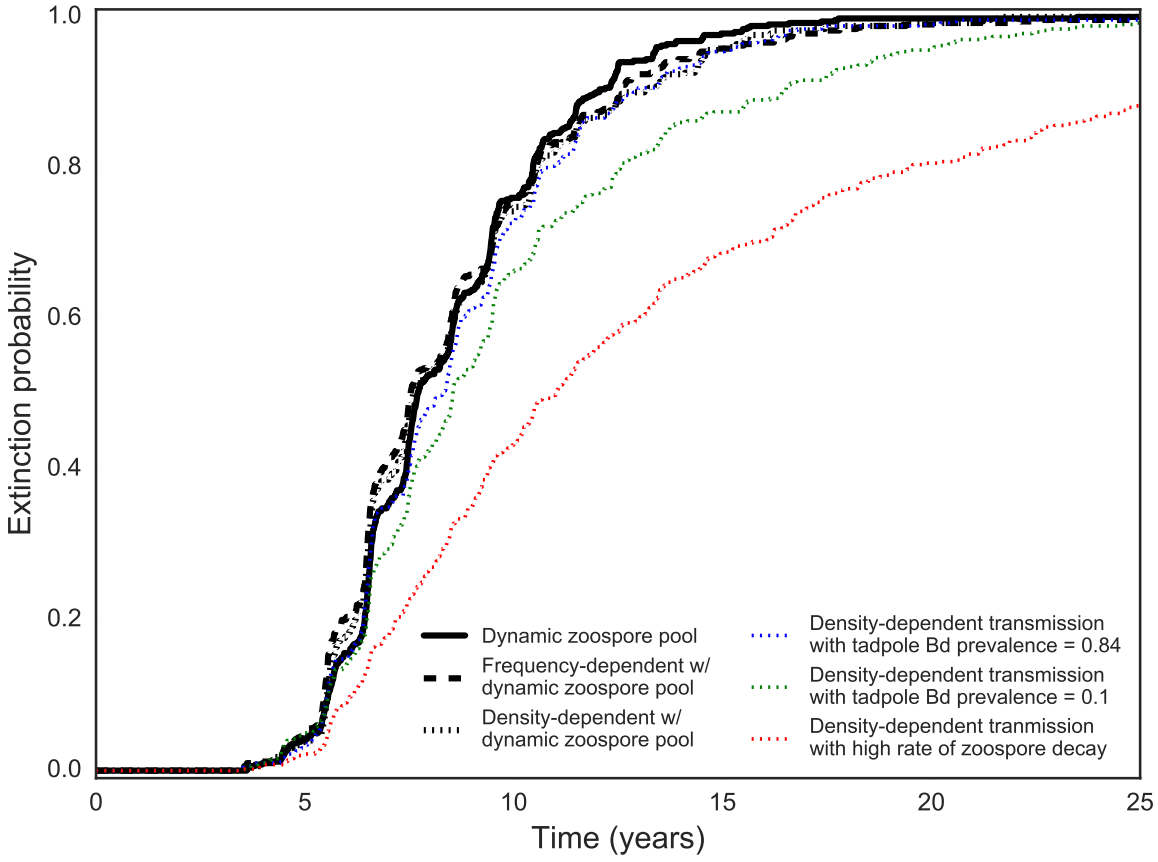


Figure 5.S5: The black lines show the time-dependent extinction curves of the hybrid model with three different transmission functions described in Table 1: dynamic zoospore pool, frequency-dependent transmission with a dynamic zoospore pool, and density-dependent transmission with a dynamic zoospore pool. The curves were generated from 500 stochastic simulations of each model parameterized with the parameters given in Table 1 and 2. The red line gives the extinction curve when transmission with density-dependent with a dynamic zoospore pool and the zoospore survival probability was set to a constant 0.05 per three day time step, compared to the laboratory-estimated temperature-dependent survival probability of 0.8 at 15 °C (Woodhams et al., 2008). The blue and green lines given the extinction curves when transmission was density-dependent with a dynamic zoospore pool and tadpole Bd prevalence was 0.84 and 0.1, respectively.

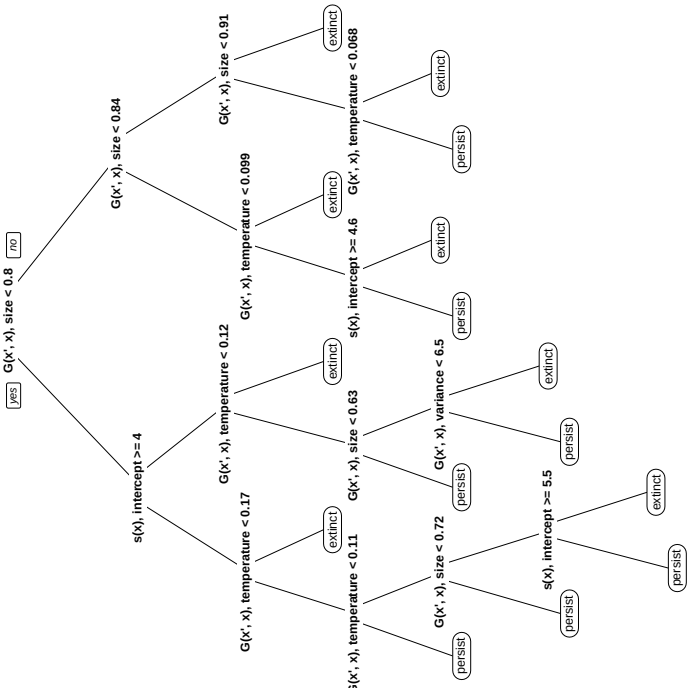


Figure 5.S6: The pruned regression tree for the model with density-dependent transmission with a dynamic zoospore pool. Shows a visual representation of how the various transmission, resistance, and tolerance parameters interacted to affect *Bd*-induced host extinction. In particular, parameters of the growth function determining host resistance ($G(x', x)$) and the survival function determining host tolerance ($s(x)$) interact to predict extinction or persistence.

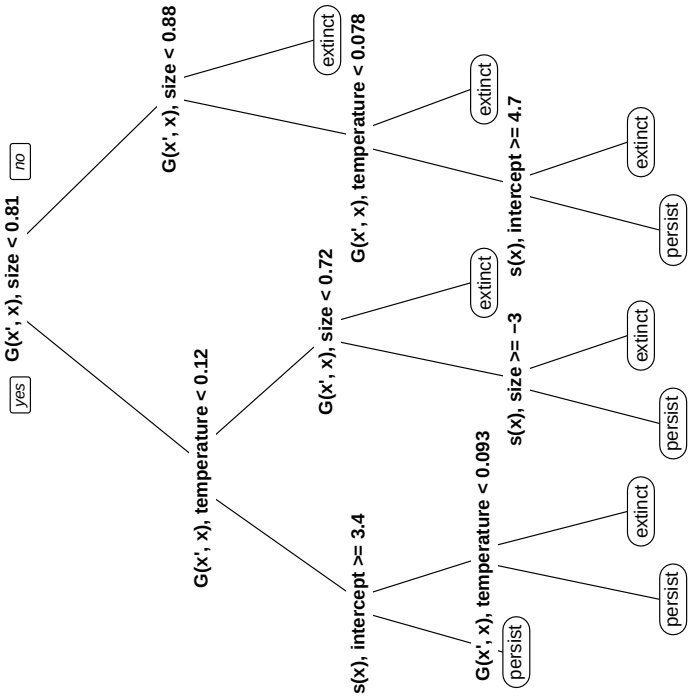


Figure 5.S7: The pruned regression tree for the model with frequency-dependent transmission with a dynamic zoospore pool. Shows a visual representation of how the various transmission, resistance, and tolerance parameters interacted to affect *Bd*-induced host extinction. In particular, parameters of the growth function determining host resistance ($G(x', x)$) and the survival function determining host tolerance ($s(x)$) interact to predict extinction or persistence.

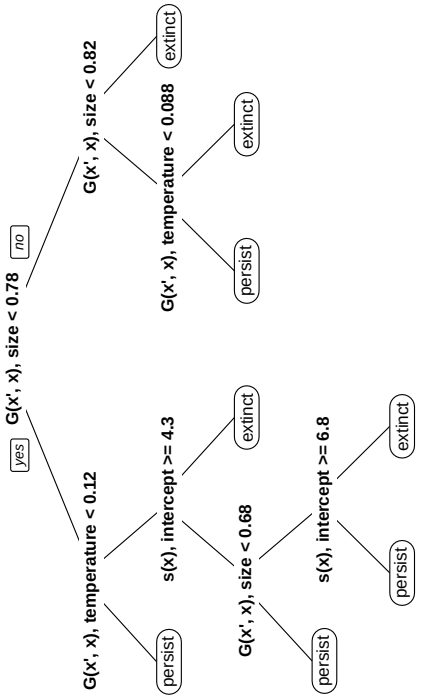


Figure 5.S8: The pruned regression tree for the model with transmission from a dynamic zoospore pool. Shows a visual representation of how the various transmission, resistance, and tolerance parameters interacted to affect *Bd*-induced host extinction. In particular, parameters of the growth function determining host resistance ($G(x', x)$) and the survival function determining host tolerance ($s(x)$) interact to predict extinction or persistence.

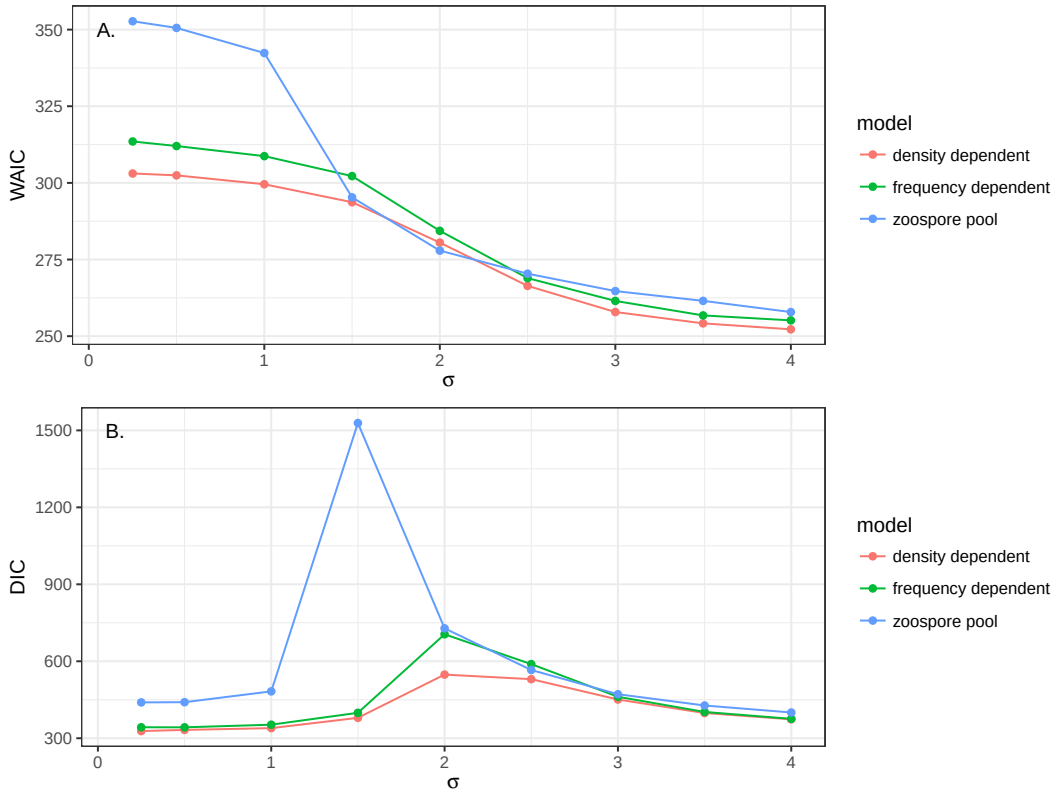


Figure 5.S9: Comparing two information criteria when fitting the latent zoospore model to the mesocosm transmission experiment with different levels of process error in the latent zoospore pool (given by σ). The above plots show WAIC (**A.**) and DIC (**B.**) values for three different transmission models fit to the experimental data: density-dependent transmission with a dynamic zoospore pool (pink), frequency-dependent transmission with a dynamic zoospore pool (green), and only dynamic zoospore pool (blue). When $\sigma \leq 1$ and $\sigma \geq 2.5$, the relative model rankings are consistent for both WAIC and DIC. However, between 1 and 2.5 the information criteria for the dynamic zoospore pool transmission function either drastically decreases (WAIC) or drastically increases (DIC).

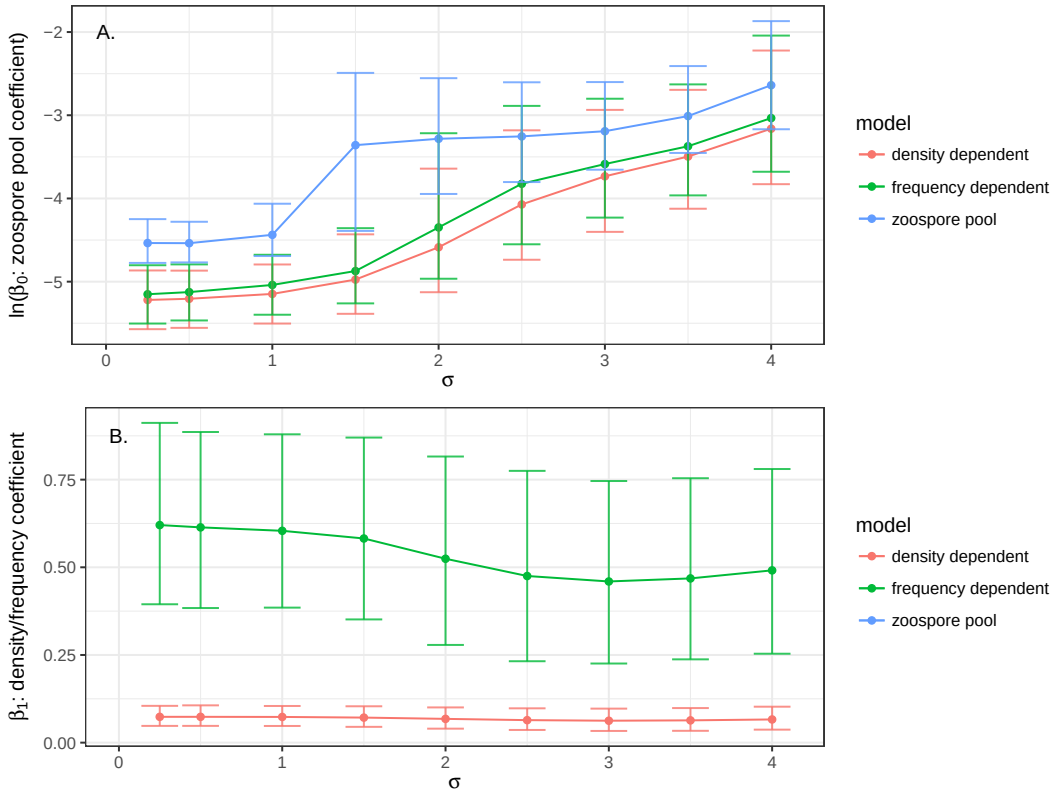


Figure 5.S10: Comparing the estimates of the transmission coefficients (β_0: the zoospore pool coefficient, β_1: the density/frequency dependent coefficient) when fitting the latent zoospore model to the mesocosm transmission experiment with different levels of process error (given by σ). **A.** The zoospore pool coefficient shows a marked increase after $\sigma = 1$. This is driven by the large variability in the trajectory of the zoospore pool, such the transmission coefficient needs to increase in order to contribute to transmission when, by chance, the zoospore pool may crash to very low levels under the high σ/process error scenarios. **B.** In contrast, the transmission coefficient β_1 determining density or frequency-dependent host contacts is relatively consistent across σ, with a slight decreasing trend. These is no blue line in B. because the transmission model with only a dynamic zoospore pool does not have a coefficient β_1. The error bars are 95% credible intervals around the coefficient estimates.