

**Tools for the occurrence of free-living and symbiotic
organisms in space and time**

by

Maxwell B. Joseph

B.S., University of California Davis, 2008

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Ecology and Evolutionary Biology
2016

This thesis entitled:
Tools for the occurrence of free-living and symbiotic organisms in space and time
written by Maxwell B. Joseph
has been approved for the Department of Ecology and Evolutionary Biology

Prof. Pieter T. J. Johnson

Prof. Vanja Dukic

Prof. Samuel Flaxman

Prof. Andrew Martin

Prof. Valerie McKenzie

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Joseph, Maxwell B. (Ph.D., Ecology and Evolutionary Biology)

Tools for the occurrence of free-living and symbiotic organisms in space and time

Thesis directed by Prof. Pieter T. J. Johnson

This doctoral dissertation broadly aims to improve methods for understanding the occurrence of organisms in space and time, including organisms that cause disease. The processes that drive occurrence are often represented mathematically as ecological theories, which can be applied to uncover actionable insights into disease management. This was a focus of my first chapter which characterized the gap between disease ecology theory and infectious disease management. My second chapter focused on filling in a gap theoretically at the intersection of multi-host pathogens and host community disassembly. Specifically, I construct a theoretical model to evaluate disease consequences when hosts are extirpated according to a variety of rules predicted from community ecology. The third chapter develops a method to link representations of latent processes that drive species occurrence to observable (with error) data, with the goal of understanding multiple causal pathways in an occupancy model. The fourth also extends occupancy models, but to allow for multi-host multi-symbiont (parasites included) systems. Chapter five is somewhat different topically, focusing on missing species trait interpolation, but methodologically is not much of a departure, drawing on Bayesian hierarchical modeling as in the previous two chapters. The final chapter develops a theoretical model to generate predictions about the effect of host diversity on the diversity and transmission of symbiotic organisms.

Dedication

To my family and friends, bipedal and otherwise.

Acknowledgements

Forthcoming

Contents

Chapter

1	Taming wildlife disease: bridging the gap between science and management	1
1.1	Why manage wildlife diseases?	2
1.2	Assessing theory application in WDM literature	4
1.2.1	Systematic search protocol	4
1.3	Results	6
1.4	Overcoming challenges to theory-based management	7
1.4.1	Bringing Together Academics and Managers: Obstacles and Opportunities .	7
1.4.2	Meeting the Needs of Managers and Academics	9
1.4.3	Making Theory Explicit Through Mathematical Modelling	10
1.4.4	Cautionary Notes and Limitations	11
1.4.5	New Approaches to Reducing Host Susceptibility: Co-infection Dynamics and Probiotics	12
1.4.6	Improving Transmission Interventions in Populations	13
1.4.7	Community Matters: Predators, Competitors and Multi-host Parasites	15
1.4.8	Evolutionary Responses to Management: A Black Box?	16
1.5	Conclusions	17
2	Does life history mediate changing disease risk when communities disassemble?	22
2.1	Introduction	22

2.2	Determinants of Extirpation Risk	24
2.3	Determinants of Host Competence	25
2.4	Covariance Between Extirpation Risk and Host Competence	26
2.5	Exploring the Consequences of Host Competence-Extirpation Risk Relationships . .	27
2.5.1	Model structure	27
2.6	Results	30
2.7	Discussion and future directions	32
2.8	Conclusions	34
2.9	Acknowledgements	35
3	Integrating occupancy models and structural equation models to understand species occurrence	44
3.1	Introduction	45
3.2	Study system	47
3.3	Data collection	47
3.4	Conceptual model	49
3.5	Model formalization	49
3.6	Estimation	55
3.7	Model assessment	55
3.8	Results	56
3.8.1	Parameter recovery	56
3.8.2	Empirical results	56
3.9	Discussion	57
3.10	Management implications	59
3.10.1	Conclusion	61
4	Multilevel models for the distribution of hosts and symbionts	69
4.1	Introduction	70

4.2	Methods	71
4.3	Case study: amphibian communities and their parasites	75
4.4	Parameter estimation	76
4.5	Results	76
4.6	Discussion	78
5	Species trait prediction via phylogenetic latent factors	86
5.1	Introduction	87
5.2	Methods	88
5.2.1	Data description	88
5.2.2	Model structure	89
5.2.3	Parameter estimation	91
5.2.4	Evaluating predictive performance	91
5.3	Results	92
5.4	Discussion	92
6	Host diversity begets symbiont diversity but reduces symbiont transmission	102
6.1	Introduction	103
6.2	Methods	105
6.3	Results	107
6.4	Discussion	109
Bibliography		121

Tables**Table**

1.1 Selected theoretical concepts in disease ecology	21
2.1 Parameter definitions	35

Figures

Figure

1.1	Management objectives reported in the wildlife disease management literature	18
1.2	Temporal trends in the application of theory, and reported management outcome distributions	19
1.3	Theoretical concepts applied in the wildlife disease literature	20
2.1	Life-history basis for covariance between extirpation risk and host competence	36
2.2	Effects of extirpation on disease risk	37
2.3	Sensitivity to interspecific transmission intensity	38
2.4	Strict allometric scaling between individual weight and intraspecific transmission rate	39
2.5	Relaxed allometric scaling between individual weight and intraspecific transmission rate	40
2.6	Effects of extirpation on disease risk with random extinctions	41
2.7	Effect of body size scaling on disease outcomes	42
2.8	Sensitivity to R_0 generative distributions	43
3.1	Conceptual model showing influences on amphibian breeding success.	62
3.2	DAG relating latent and observed variables.	62
3.3	Direct effects on amphibian occurrence	63
3.4	Submodel fit for observed indicator variables	64
3.5	Parameter recovery and interval coverage	65

3.6 Recovery of species-specific effects	65
3.7 Among refuge variance parameters for occupancy probabilities	66
3.8 Detection parameter estimates	67
3.9 Graphical independence checks	68
4.1 Site level variance covariance matrix and random effect posteriors	81
4.2 Survey level variance covariance matrix and random effect posteriors	82
4.3 Host species level variance covariance matrix and random effect posteriors	83
4.4 Individual level variance covariance matrix and random effect posteriors	84
4.5 Bivariate posterior distributions of the effective variance and dependence	85
5.1 Missing trait map matched to the species level phylogeny	95
5.2 Correlation functions considered for the latent factors	96
5.3 Performance curves for models with varying phylogenetic correlation structures . . .	97
5.4 Bivariate posterior distribution for the correlation between held-out values and posterior expected values	98
5.5 Observed (held out) values and model predictions	99
5.6 Mean squared predictive error as a function of the number of trait observations . .	100
5.7 Posterior predictions with 95% prediction intervals, separated by trait	101
6.1 Niche space for symbionts	114
6.2 Predictions when niche widths are equal	115
6.3 Predictions when niche widths vary among communities	116
6.4 Predictions when fitness consequences vary among communities	117
6.5 Predictions when niche breadth varies within communities	118
6.6 Predictions when fitness consequences vary within communities	119
6.7 Predictions when communities include symbionts with varying niche widths and fitness consequences	120

Chapter 1

Taming wildlife disease: bridging the gap between science and management

MAXWELL B. JOSEPH, JOSEPH R. MIHALJEVIC, ANA LISETTE ARELLANO, JORDAN G. KUENEMAN, DANIEL L. PRESTON, PAUL C. CROSS, AND PIETER T. J. JOHNSON

Parasites and pathogens of wildlife can threaten biodiversity, infect humans and domestic animals, and cause significant economic losses, providing incentives to manage wildlife diseases. Recent insights from disease ecology have helped transform our understanding of infectious disease dynamics and yielded new strategies to better manage wildlife diseases. Simultaneously, wildlife disease management (WDM) presents opportunities for large-scale empirical tests of disease ecology theory in diverse natural systems. To assess whether the potential complementarity between WDM and disease ecology theory has been realized, we evaluate the extent to which specific concepts in disease ecology theory have been explicitly applied in peer-reviewed WDM literature. While only half of WDM articles published in the past decade incorporated disease ecology theory, theory has been incorporated with increasing frequency over the past 40 years. Contrary to expectations, articles authored by academics were no more likely to apply disease ecology theory, but articles that explain unsuccessful management often do so in terms of theory. Some theoretical concepts such as density-dependent transmission have been commonly applied, whereas emerging concepts such as pathogen evolutionary responses to management, biodiversity-disease relationships and within-host parasite interactions have not yet been fully integrated as management considerations. Synthesis and applications. Theory-based disease management can meet the needs of both academics and managers by testing disease ecology theory and improving disease interventions. Theoretical con-

cepts that have received limited attention to date in wildlife disease management could provide a basis for improving management and advancing disease ecology in the future.

1.1 Why manage wildlife diseases?

Approximately 43In recent decades, a growing acknowledgement of the role of wildlife diseases in human health has prompted research on pathogens that have directly or indirectly jumped from wildlife to humans, such as human immunodeficiency virus (HIV), severe acute respiratory syndrome (SARS) virus, West Nile virus and various influenza viruses resulting in a rapid increase in the understanding of disease ecology (Hudson et al., 2002; Ostfeld et al., 2010; Wobeser, 2013). Further impetus for disease ecology research stems from wildlife diseases that afflict domestic animals and cause economic losses due to direct mortality, farm-wide culling and trade restrictions, as seen with classical swine fever, foot-and-mouth disease, bovine tuberculosis and brucellosis (Keeling et al., 2001; Schnyder et al., 2002; Woodroffe et al., 2006; Cross et al., 2007).

In addition to impacts on human health and economies, wildlife diseases are increasingly recognized as a conservation challenge. Formerly, it was widely thought that pathogens would not cause extinction because as host population density decreased due to disease, contact rates would become too low for transmission to continue; thus, pathogens would be extirpated before host populations (May and Anderson, 1979). However, disease-induced declines and extinctions of wildlife resulting from small population sizes, reservoir hosts, host switching and heterogeneity in contact rates, susceptibility and transmission within and among populations have forced a re-evaluation of this perspective (de Castro and Bolker, 2004). For example, when contact rates among individuals do not depend on host density, pathogens are more likely to drive populations to extinction because transmission continues as host populations are reduced, as seen with Tasmanian devil *Sarcophilus harrisii* (Boitard 1841) facial tumour disease where transmission appears to be related to mating behaviours (McCallum, 2012). White nose syndrome in bats also seems to be more likely to cause extinctions owing to social behaviours in which hosts cluster in hibernacula, reducing the correlation between contact rates and population densities (Langwig et al., 2012).

In recent decades, wildlife disease management (WDM) has been increasingly used to conserve threatened wildlife populations (Deem et al., 2001). For example, WDM has controlled outbreaks of feline leukaemia in critically endangered Iberian lynxes *Lynx pardinus* (Temminck 1827) and rabies in endangered Ethiopian wolves *Canis simensis* (Rppell 1840), both of which are associated with domestic animal disease reservoirs (Haydon et al., 2006; López et al., 2009).

Despite the wealth of empirical WDM research, management outcomes can be difficult to predict because system-specific information is lacking for novel pathogens and many theoretical concepts in disease ecology (see Table 1 for a subset) have not been widely tested in the field, leading to uncertainty in their generality. This is unlike other environmental management disciplines such as fisheries ecology, which has effectively used theoretical models to predict yields, manage harvest timing and limits and design reserves (Barbara et al., 2003). Indeed, theoretical applications in fisheries ecology have also produced insights into density-dependent population dynamics, metapopulation theory and the evolution of life-history strategies (Frank and Leggett, 1994). In this review, we assess the extent to which a similar union between theory and practice has been achieved in WDM.

We use a quantitative, case-based approach to provide a critical retrospective of WDM over the last four decades to: (i) quantify how frequently specific theoretical concepts from disease ecology have been applied in the literature, (ii) identify prevailing management objectives, groups and reported outcomes and (iii) assess taxonomic biases in WDM literature. We then present methodological and conceptual opportunities to facilitate the newly emerging synthesis of disease ecology and management, drawing from environmental management and biomedicine to outline steps towards more cost-effective, efficacious and informative WDM. This synthesis aims to facilitate the development of a more predictive framework for disease interventions while simultaneously building empirical support for understanding of disease processes across systems.

1.2 Assessing theory application in WDM literature

1.2.1 Systematic search protocol

We compiled WDM case studies using a systematic, two-step search process with specific criteria for inclusion in our review. In the first stage, we searched titles and abstracts of records included in ISI Web of Science using specific terms [(wild*) AND (disease* OR infect* OR pathogen* OR parasit*) AND (manage* OR conserv*)] to capture breadth in published WDM records. Additionally, we searched for case studies in grey literature using the following online resources: National Wildlife Health Center, Wildpro, National Biological Information Infrastructure Wildlife Disease Information Node, U.S. Government Printing Office and the U. S. Fish and Wildlife Service. No case studies identified in the grey literature met our criteria that were independent of cases identified in the scientific literature. Case studies were also identified using previous review papers and books (Lafferty and Gerber, 2002; Wobeser, 2002; Hudson et al., 2002; Keesing et al., 2006; Wobeser, 2013).

We conducted a follow-up search with ISI Web of Knowledge to capture subject depth for each managed disease or pathogen identified in the first step, using a search string that included all pathogen and disease names along with terms related to management interventions: (e.g. (rabi* OR lyssavir*) AND (vaccin* OR treat* OR manag* OR control* OR preval* OR incidence OR cull*) AND (wild* OR free-ranging OR free ranging)). The initial Web of Science search returned articles dating back to 1989, but our disease-specific search strings often returned results dating back to the 1950s or earlier. Historical accounts of WDM are probably under-represented in the literature available online, and those returned by our search strings were often less readily accessible than recently published articles. As a result, the cases reviewed here primarily represent recently published cases of WDM. The publication dates of included cases range from 1973 to 2011, and 75

For each article that met our criteria, we recorded (i) pathogen and host characteristics, (ii) management motivations, strategies and outcomes and (iii) whether and how disease ecology theory was incorporated in each article that satisfied our criteria. We only included cases that

provided quantitative data on disease in a population or area (number of cases, seroprevalence, prevalence, incidence, etc.). When multiple records were encountered for a single management event, we used the most recent record (as of Spring 2011). Cases that only described disease management in humans, livestock or plants were excluded. Finally, we only included studies that described management of diseases in populations (operationally defined as groups of ≥ 1 individual) of free-ranging wildlife.

Incorporation of disease ecology theory was defined broadly as the explicit use or discussion of theoretical concepts relating to transmission dynamics, host population regulation by disease, pathogen evolution, host or pathogen community effects on transmission, spatial heterogeneity in disease dynamics, life stage- or age-specific disease dynamics, endemic vs. epidemic disease states and herd immunity (see Table 1 for a list of specific concepts used to define theory in the literature search).

Four broad management objectives were identified, including conservation of a host species, prevention of disease transmission to humans, prevention of disease transmission to livestock and basic research. Studies falling into our basic research category were usually an attempt to better understand the system, determine the extent of the disease problem or provide insight into future management opportunities. To investigate differences in theory application and objectives among managing groups, we also classified author affiliations for each paper as academic, governmental, private or some combination thereof. University or university laboratory affiliations were considered academic, and we used the same criteria for governmental and private affiliation. Mixed author affiliations (e.g. academic and governmental) were recorded for individual authors and for papers with multiple authors with different affiliations.

We characterized management outcomes by recording whether the disease was eradicated, and if not, whether there were changes in the prevalence, incidence or intensity of disease. Ideally, these changes could be quantified and compared across disease systems, but in many cases, inconsistent reporting of results and a lack of pre-management or control data complicate meaningful quantitative comparisons of effect sizes across studies. Finally, we considered whether the origi-

nal management objective was attained using the following categories: apparent success, meaning that there was no unmanaged control population or area to compare to the treated area; partial success, meaning that at least some of the management objectives were reported as fulfilled; and success, for cases that had controls and reported fulfillment of all management objectives. While management outcomes are rarely clear-cut in this practice, this simplified classification system facilitated coarse comparisons across disease systems and among management studies with variable monitoring time-scales.

1.3 Results

In total, 101 scientific articles among the 14 275 identified from the search strings satisfied our criteria. Many (40) conservation motivated 87 overall, host conservation was the most common objective (39).

Disease ecology theory as defined above has only recently been incorporated consistently into WDM literature (Figure 1.2). Some theoretical concepts such as density dependence in transmission were frequently applied, while others such as pathogen evolution and the role of predators and biodiversity in regulating disease were not (Figure 1.3, Table 1). Unexpectedly, papers authored by academics were not more likely to incorporate theory (Fisher's exact test, $P = 0.909$). Management outcomes were related to theory incorporation (Fisher's exact test, $P = 0.042$). The three papers that reported disease increases following intervention explained their results in terms of disease ecology theory, providing insights into transmission and optimal control strategies (Figure 1.2). However, there was no relationship between management objective attainment and theory incorporation (Fisher's exact test, $P = 0.746$). Nevertheless, some counter-intuitive but successful management programmes clearly benefited from theory. For example, control of classical swine fever in wild boar *Sus scrofa* (Linnaeus 1758), is often hampered by stage-dependent transmission dynamics. Susceptible piglets are hard to target with baited vaccines and act as disease reservoirs. By allowing an epidemic to peak such that most adults are immune, then culling only piglets, Swiss academics and governmental groups successfully eradicated the disease from a 166-km² region near

the Italian border (Schnyder et al., 2002).

Reductions in prevalence, incidence or infection intensity were reported in 75Ninety-four percent of cases reported management in terrestrial systems, with 4The majority (89However, mammals are less speciose and less threatened by disease than amphibians (Vié et al., 2009), for which we found no published WDM records. Taxonomic bias could arise because vaccines and drugs are developed primarily to protect human, livestock or poultry health. Relatively few cases (13

Collectively, our analyses indicate that while academics and government agencies collaborate to manage wildlife diseases, collaborations do not necessarily lead to an integration of disease ecology theory with management. Density-dependent transmission was often assumed to justify control efforts, but other theoretical concepts were rarely applied (Figure 1.3). Data quality issues and potential publication biases currently hinder the application of meta-analytical techniques for WDM, and there is a paucity of published records on non-mammalian management.

1.4 Overcoming challenges to theory-based management

1.4.1 Bringing Together Academics and Managers: Obstacles and Opportunities

While collaboration alone may not necessarily lead to an integration of disease ecology theory and WDM, it should provide a starting point for such integration. Academics and managers have unique needs, constraints and knowledge-seeking behaviour that challenge such collaborations. For instance, untreated control areas or pre-treatment data can be unavailable or even unethical in WDM, but are critical for experiments in disease ecology. While academics may design field experiments to test and refine theoretical models, managers need practical, effective and uncontroversial management strategies that succeed in particular systems. Such strategies may not be easily identified in the literature from model systems, which managers may be unable to access.

Modelling wildlife disease systems requires decisions about model complexity. In our experience, theoreticians prefer simpler, more general models that may be applicable to many systems.

These models are easier to parameterize and analyse, and the resulting papers are likely to have a wider academic audience. On the other hand, simple models are easily discarded by managers because they lack system-specific detail. This tension is likely to continue, but we recommend additional flexibility on both sides. In particular, managers should appreciate that the addition of modelling details that are only weakly supported by data may not lead to better predictions. Meanwhile, theoreticians may develop general models that bear little resemblance to any biological system. Furthermore, individuals may be most interested in a particular suite of theoretical concepts, but a narrow approach can impede management by ignoring the full range of phenomena relevant to producing desired management outcomes (Driscoll and Lindenmayer, 2012). Thus, academics and managers are challenged to take a broad view that incorporates relevant theoretical concepts and an appropriate amount of biological realism, which may require collaboration among researchers with different areas of expertise (Driscoll and Lindenmayer, 2012). Unfortunately, such large collaborative efforts may bring a loss of autonomy at odds with academic or governmental bureaucracy.

A diverse body of literature addresses the gap between academics and environmental managers and provides examples of successful strategies for integration. For instance, international symposia have improved information transfer in invasion biology (Shaw et al., 2010). Social networking, joint appointments, interinstitutional sabbaticals, fellowships, concise reporting of relevant science to managers and targeted calls for research proposals by managers can all help to foster cooperation (Gibbons et al., 2008). Interdisciplinary working groups for particular management issues can ensure that the needs of multiple stakeholders are considered together when organizing such activities (Gibbons et al., 2008). Groups such as the Wildlife Disease Association and applied journals including the Journal of Wildlife Diseases have encouraged interdisciplinary collaboration, and a broader recognition of the complementarity between disease ecology theory and WDM can provide the impetus for expanding interdisciplinary work in this important field.

1.4.2 Meeting the Needs of Managers and Academics

Theory can help address unprecedented management challenges and can be refined in the process. Disease outbreaks are often caused by novel pathogens or the appearance of known pathogens in new hosts. Often, details of hostpathogen interactions are unknown. By combining limited information with general principles of disease ecology (Table 1), management actions can be taken and research priorities identified, informing management as data accumulate (McDonald-Madden et al., 2010). Management can test theoretical predictions using large-scale manipulations in real systems, often logistically unfeasible for academics, producing insights and publications even in a crisis. Field interventions can test concepts of unknown generality in wildlife disease systems. This is particularly important for management because interventions such as culling are justified on the basis of host density thresholds for disease persistence, which depend on host life-history, seasonality and population dynamics (Lloyd-Smith et al. (2005); Altizer et al. (2006); Figure 1.3).

Theory is often refined by evaluating competing hypotheses. Therefore, adaptive management is one way to integrate theory and management, especially if multiple management hypotheses can be tested (Holling and Others, 1978). Differentiation among competing hypotheses is synonymous with identifying optimal management in this framework. Thus, monitoring the effects of disease interventions on prevalence, virulence and host vital rates can help to estimate model parameters including transmission and recovery rates and help in evaluating management outcomes. When agencies have limited flexibility in decision-making, thus precluding adaptive management, the best available theoretical and system-specific knowledge can at least produce a best guess management strategy (Gregory et al., 2006). Failed management is still valuable in this framework because outcomes can be compared to predictions from competing models of disease dynamics that can be selectively eliminated, as with Tasmanian devil facial tumour disease (McDonald-Madden et al., 2010). This approach produces mechanistic insights that might be missed if management strategies are characterized simply as effective or ineffective based on management outcomes.

If many groups apply adaptive management separately in similar systems without commun-

cating, generalities that benefit management and disease ecology may remain elusive. Systematic reviews, invaluable to biomedicine, can help establish which interventions are effective and explain heterogeneity in effectiveness with a standardized meta-analytic approach. Guidelines for systematic reviews in environmental management exist, but have not been applied in WDM (Pullin and Stewart, 2006). Our metadata indicate that this may be due to a lack of data quality and quantity.

Simple recommendations to facilitate the production of data suitable in quality for systematic review include: (i) establishing unmanaged control areas and/or baseline data, (ii) achieving replication, (iii) reporting precision for estimates of model parameters, prevalence and effect size, (iv) publishing and mechanistically explaining failed management and (v) reporting the spatiotemporal extent of management. Data quantity may be lacking because of publication biases and a lack of incentives for managers to publish when working independently. This latter issue is minor if collaborations involve academics, but even motivated scientists may have difficulty publishing if management has no effect. However, management failures are as important to report in the literature as successes for systematic reviews and meta-analyses. An evaluation of the applicability of a theoretical concept in a particular case will rely on comparisons of observed data with explicit predictions from theoretical models, which can often be derived through mathematical modelling.

1.4.3 Making Theory Explicit Through Mathematical Modelling

Theoretical concepts can be explicitly built into system-specific mathematical models to identify and evaluate management strategies, as exemplified in a modern WDM challenge: chronic wasting disease (CWD). In 2003, the state of Wisconsin began culling white-tailed deer *Odocoileus virginianus* (Zimmermann 1780) and lengthened the hunting season in an attempt to eradicate CWD. These efforts were mandated despite uncertainty over transmission dynamics, the environmental persistence of prions that cause CWD and the time of CWD arrival to the state (Bartelt et al., 2003). Five years later, prevalence was still slowly increasing (Heisey et al., 2010). As this epidemic has unfolded, several models have been used to describe the dynamics of CWD (Gross and Miller, 2001; Wasserberg et al., 2009; Wild et al., 2011). Simple models of CWD do not tend

to produce plausible results. Purely density-dependent transmission models predict increases in prevalence that are too rapid, while frequency-dependent transmission models predict rapid host extinction or epidemics that are very slow to develop (on the order of centuries).

Modelling indirect transmission via environmental contamination results in a wider range of outcomes and produces several patterns observed in the field including a slow disease progression with prevalences of over 30%. Recent analyses did not provide much support for models with variable increases in transmission over models with variable starting prevalence, suggesting that host density effects may be relatively weak in this system (Heisey et al., 2010). Taken together, these results suggest that managers would have to reduce deer densities to extremely low levels, probably for decades, at which point other stakeholders such as hunters may wonder whether it is worse to have a lower deer density due to disease impacts or disease control efforts.

1.4.4 Cautionary Notes and Limitations

Disease ecology theory is not a silver bullet for solving management problems. Indeed, some have pointed out that application of theory under certain circumstances can lead to poor management (Driscoll and Lindenmayer, 2012). Misapplication of theory at an inappropriate scale, or in a system that does not meet necessary assumptions, could cause undesired consequences. For instance, an assumption of broad-scale culling as a disease management intervention is that pathogen transmissions scale positively with host population density. However, density-dependent changes in social behaviour can alter dispersal patterns that violate this assumption, increasing transmission, as seen with bovine tuberculosis (TB) in cattle and European badgers *Meles meles* (Linnaeus 1758) (Woodroffe et al., 2006). Work in the badger-TB system has refined our understanding of the effects of culling on social animals. However, one could argue that if culling-induced dispersal had been discovered in another disease system, the unintended increase in TB transmission to cattle following badger culling might have been avoided. Omissions involve academics, but even motivated scientists may have difficulty publishing if management has no effect. However, management failures are as important to report in the literature as successes for systematic reviews and meta-analyses. An

evaluation of the applicability of a theoretical concept in a particular case will rely on comparisons of observed data with explicit predictions from theoretical models, which can often be derived through mathematical modelling.

1.4.5 New Approaches to Reducing Host Susceptibility: Co-infection Dynamics and Probiotics

Recent advances in disease ecology based on co-infection provide new ways to reduce disease susceptibility and transmission. For example, in African buffalo *Syncerus caffer* (Sparrman 1779), gastrointestinal nematodes reduce individual resistance to *Mycobacterium tuberculosis*, which causes bovine TB, because of cross-regulatory immune responses to micro- and macroparasites (Ezenwa et al., 2010). Hence, deworming drugs may increase resistance to TB and improve TB vaccination efficacy, raising the possibility that TB could be managed indirectly through nematode control (Elias et al., 2006; Ezenwa et al., 2010).

Management involving immunological trade-offs could improve general understanding of immune-mediated parasite interactions. For instance, interventions aimed at helminth parasites, which accounted for 25These predictions could be evaluated opportunistically by monitoring non-target pathogens. Similarly, management in systems with co-infecting parasites could be used to understand virulence evolution in response to changing co-infection dynamics (Alizon and van Baalen, 2008).

There is increasing recognition that microbial symbiosis can play a role in host health. Using mutualistic microbes to control disease, a technique known as probiotics therapy, has benefitted aquaculture, agriculture and human medicine. For example, addition of *Bacillus* and *Pseudomonas* bacteria controls pathogenic Vibrio that infect prawns, salmon and crabs in aquaculture (Irianto and Austin, 2002; Panigrahi and Azad, 2007). *Bifidobacterium* and *Lactobacillus* can ameliorate *Escherichia coli* infection in pig farms (Zani et al., 1998; Shu et al., 2001). In humans, probiotics can treat diarrhoea caused by *Clostridium difficile* infection and antibiotic therapy (McFarland, 2006; Rohde et al., 2009).

Probiotics may prove useful for WDM. Frogs with certain skin bacteria may be less susceptible to population extirpation caused by chytridiomycosis, a fungal disease that implicated global amphibian declines (Lam et al., 2010). Experimental augmentation of skin bacteria reduces mortality of susceptible amphibians in captivity, and field experiments are underway to determine whether probiotics can prevent extirpations in nature (Harris et al., 2009). Probiotics can also reduce vector populations. For instance, laboratory-reared mosquitoes with a maternally heritable probiotic that disrupts dengue fever virus transmission can locally replace wild mosquito populations and reduce dengue fever risk (Hoffmann et al., 2011).

The successful use of probiotics depends on an understanding of microbial ecology, especially with respect to long-term probiotic maintenance in a host or environment. Risks associated with introducing non-native microbes may be ameliorated by isolating probiotic agents from native hosts. As data accumulate, it will be important to evaluate whether the risks of probiotics outweigh those associated with antibiotic treatment in terms of antibiotic or probiotic resistance and pathogen virulence evolution. Finally, linking these within-host processes to among-host processes (e.g. microbial community structure and transmission) is an important frontier for WDM and disease ecology.

1.4.6 Improving Transmission Interventions in Populations

Optimal management strategies depend on the degree to which transmission is driven by host population density and the amount of individual and population heterogeneity in contact or transmission rates. Host population size, aggregation patterns and contact rates can be altered through hunting, artificial feeding, predator and scavenger conservation, fertility control, culling, translocation of individuals, artificial stocking, movement barriers, etc. Understanding the functional form of the relationships among host contacts, density and transmission in real systems is critical to predicting the impacts of such interventions. Therefore, field manipulations will play a crucial role in refining our mechanistic understanding of disease transmission.

Optimal management strategies are not static; contact rates, host abundance and demogra-

phy can change naturally over time, in response to disease and due to management. For example, group sizes and contact rates may remain constant for highly social species despite management-induced population reduction. Reservoir hosts may increase disease risk for other species if infected individuals maintain high fitness via increased reproductive output - fecundity compensation, for example (Schwanz, 2008). Fertility control of such reservoir hosts may protect other species that are less tolerant to infection. Lastly, if transmission peaks in a short time period, perhaps due to breeding or a pulsed influx of juveniles (Altizer et al., 2006), management may be applied optimally in a narrow time interval.

Brucellosis in the elk (*Cervus elaphus* Linnaeus 1758) populations of the Greater Yellowstone Ecosystem of Wyoming illustrates how management can capitalize on temporal transmission dynamics. Every year, wildlife managers provide supplemental feed to elk at 22 sites in the region. Contrary to theoretical predictions, elk abundance at each feeding site is uncorrelated with brucellosis seroprevalence (Cross et al., 2007), but locally, host contact rates correlate positively with elk density (Creech et al., 2012). These seemingly contradictory findings are explained by variation and interaction between transmission and host density over time, which suggests that brucellosis seroprevalence may be reduced by shortening the length of the feeding season in early spring when transmission is highest (Maichak et al., 2009). This option is appealing because vaccination has had limited, if any, effect (Cross et al., 2007), and a test-and-remove programme, although effective, is financially prohibitive to implement across a broad region.

Establishing contact networks for a variety of disease systems across a range of densities and over time will help to identify life-history traits, social structures and other species characteristics that predictably influence transmission. Taken together, these population-level tools can advance general understanding of transmission dynamics and optimize the application of disease control strategies.

1.4.7 Community Matters: Predators, Competitors and Multi-host Parasites

Community-level interactions including predation and competition can influence disease management outcomes. Predation on hosts can increase or decrease disease prevalence and the likelihood of epidemics depending upon predator selectivity, as well as behavioural and demographic effects on host populations that influence transmission and disease susceptibility (Packer et al., 2003; Holt and Roy, 2007). Interspecific competition can also influence host background mortality and thus the net effect of disease in a population (Bowers and Turner, 1997). Unintended consequences when managing predators or competitors may be of less concern if coupled with ongoing management such as predator restoration and invasive species control.

Interspecific transmission of generalist parasites is hard to quantify, but attempts to control generalist parasites in one host species can reveal the extent to which other hosts contribute to transmission. For example, vaccinated white-footed mice *Peromyscus leucopus* (Rafinesque 1818) in southern Connecticut reduced the prevalence of *Borrelia burgdorferi*, the bacterium that causes Lyme disease (Tsao et al., 2004). Based on the strains of *B. burgdorferi* found in ticks in vaccinated plots, and the relationships between mouse density and tick infection prevalence, the authors concluded that other host species contributed more to tick *B. burgdorferi* infections than previously thought. Thus, vaccination would have to target multiple host species to be effective. Contact prevention between wildlife and livestock also provides an opportunity to prevent disease spillover, and when linked with monitoring of both wildlife and domestic populations, can be used to estimate relative rates of within- and among-species transmission.

Novel management strategies may target ultimate causes of disease emergence once they have been identified. For instance, Lyme disease risk in the north-eastern United States increases with habitat fragmentation, which leads to extirpations of (i) predators and competitors that limit white-footed mouse abundance and (ii) less-competent hosts for *B. burgdorferi* and ticks (Ostfeld and LoGiudice, 2003). In this system, biodiversity conservation might be an option for proactive WDM. Management interventions that recognize and target community- or ecosystem-level processes are

rare, but in some cases may more directly address disease threats than focusing solely on individuals or populations.

1.4.8 Evolutionary Responses to Management: A Black Box?

Common WDM interventions have evolutionary consequences that remain largely unexplored. In contrast, a vast literature in the biomedical sciences describes the effects of vaccination on the evolution of human pathogens. Generally, (i) some pathogens tend to evolve vaccine resistance, (ii) imperfect vaccines that confer partial immunity select for increased virulence, and (iii) live attenuated vaccines can revert to virulence if inadequately distributed (Anderson et al., 1987; Gandon and Day, 2007; Mackinnon et al., 2008). Together, these observations provide strong incentives for an all or nothing approach to vaccine-laden bait distribution programmes, which may jeopardize long-term success if low-coverage field trials using vaccines of limited or unknown efficacy precede full distribution of an effective vaccine.

Selective culling (analogous to selective predation) whereby managers remove infected individuals from the population to prevent disease spread may also have unintended consequences. It can select for increased virulence, because there are relatively more susceptible hosts available for the pathogen, and pathogens must transmit to susceptible hosts faster to avoid being culled along with their hosts (Choo et al., 2003). In many cases, selective culls are based on serological tests that do not discriminate between recovered and infectious individuals. Removal of recovered individuals may actually result in more explosive epidemics later on due to a reduction in herd immunity (Ebinger et al., 2011).

Experiments and genetic analyses of wildlife pathogens that are often treated by vaccination or culling could reveal the extent to which these concerns are realized. Aside from developing new vaccines, these risks may be mitigated if management capitalizes on immune-mediated parasite interactions, employ probiotic approaches and consider population- and community-level management interventions. The use of multiple strategies (seen in 10

1.5 Conclusions

A more complete integration of disease ecology with WDM can benefit both disciplines. Management provides unique opportunities to test disease ecology theory while building system-specific understanding. By evaluating management outcomes in terms of theory, managers can better identify effective strategies even in the face of management failures. We have presented specific recommendations, methodological tools and conceptual approaches to achieve a stronger integration of theory and practice, which we hope will facilitate the development of a strong predictive framework for WDM. The generality of this framework is currently limited by the lack of theoretical and taxonomic breadth of coverage. However, these biases are beginning to be addressed, and disease ecology theory is being integrated with WDM with increasing frequency. By continuing to incorporate ecological and evolutionary ideas in the development and evaluation of management actions, disease ecology and WDM are likely to continue to advance towards a more unified body of theory and evidence.

Figure 1.1: Distribution of management objectives across managing groups (G = government agency, A = academic researchers, P = private group, and letter combinations indicate collaborations between groups).

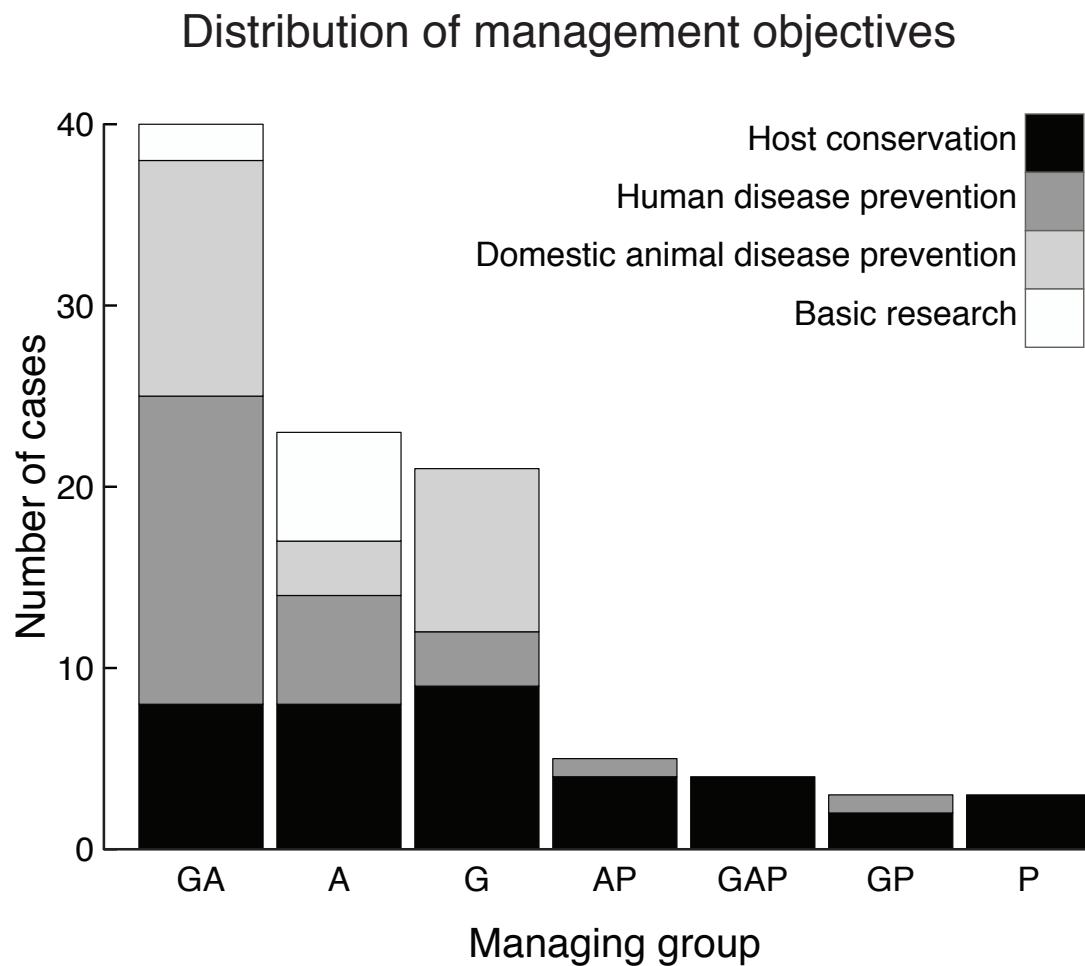


Figure 1.2: (a) Time series of theory incorporation in published WDM cases. The size of the dot is proportional to the number of cases included in our review from each time interval. (b) Distribution of management outcomes according to whether disease ecology theory was incorporated. Reductions and increases refer to changes in prevalence, incidence, infection intensity or disease-induced mortality; eradication refers to local rather than global eradication.

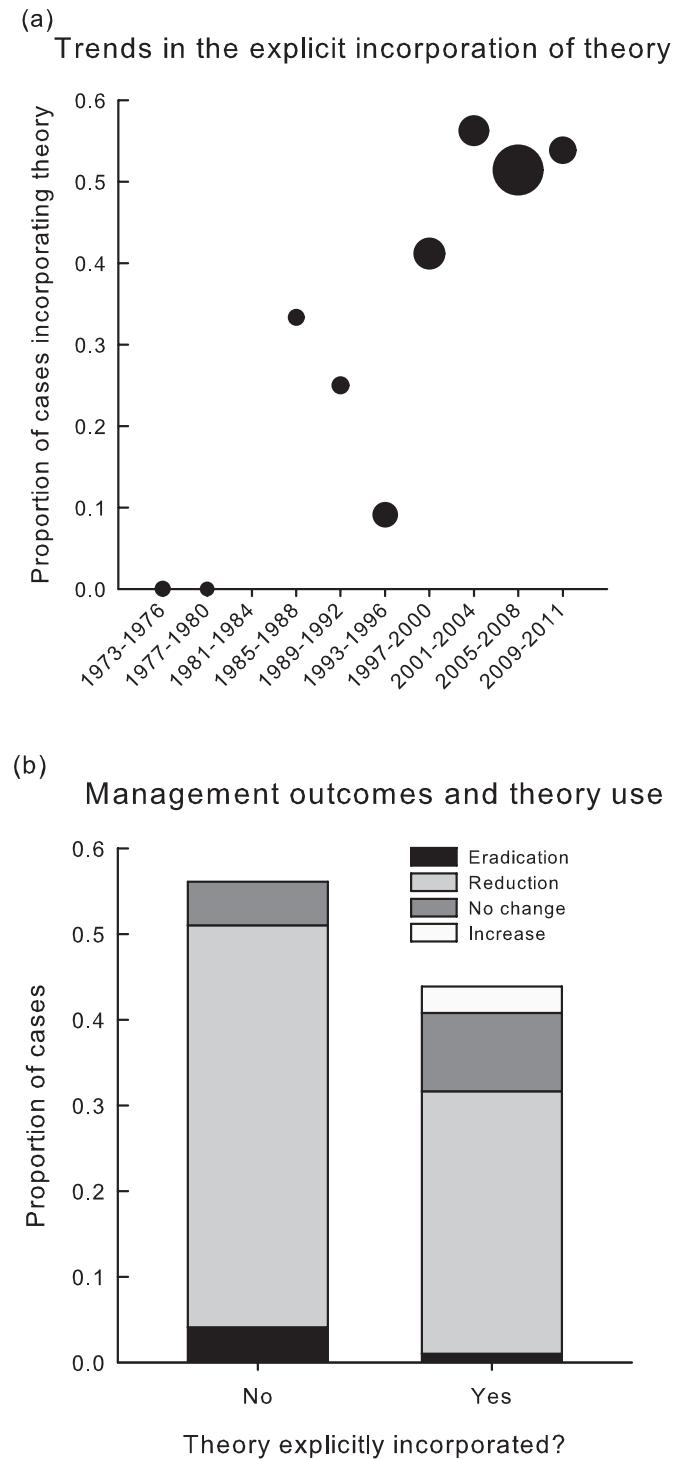


Figure 1.3: Bar plot of the specific theoretical concepts identified in Table 1, and their application in the literature was included in this review, showing that some concepts such as density-dependent transmission are well represented, while others were less frequently (or not at all) applied.

Theory applied in reviewed WDM literature

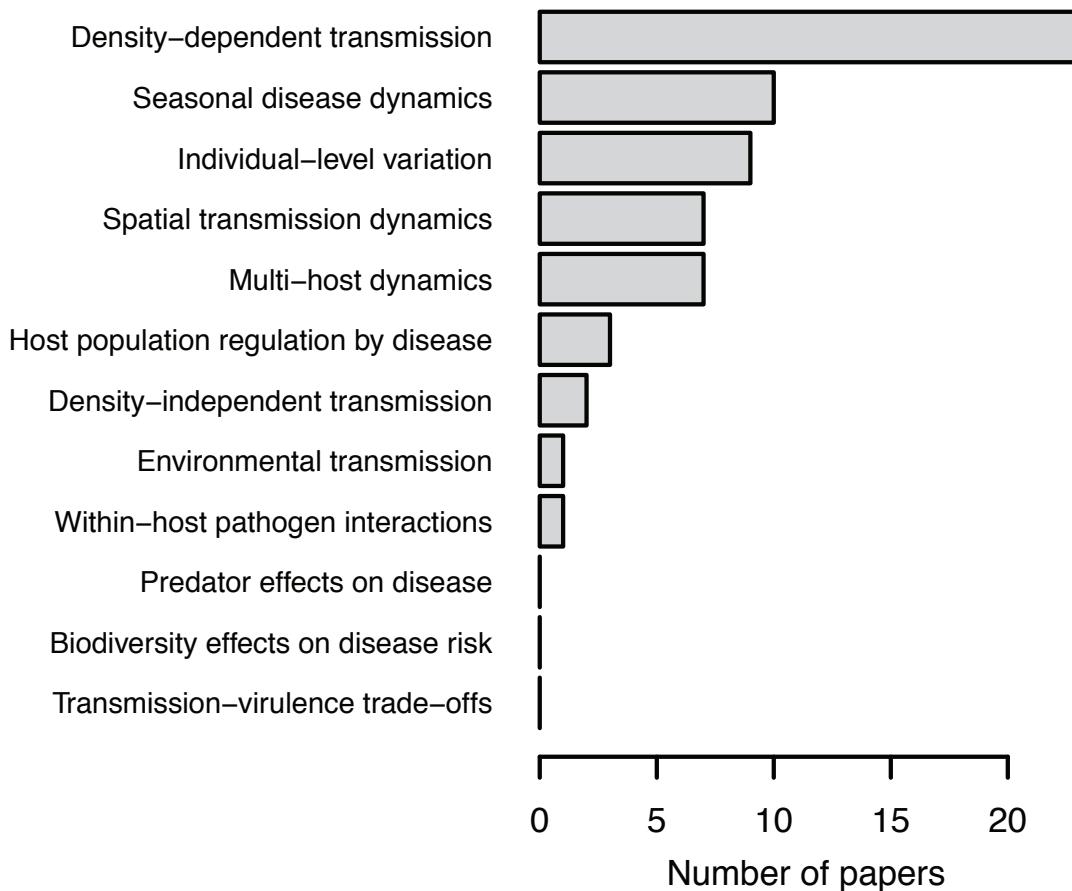


Table 1.1: Selected theoretical concepts in disease ecology

Theoretical concepts	Management applications	Refs.
Host population regulation by disease	Disease reductions may increase host abundance and/or survival	Anderson and May (1978)
Trade-offs between transmission and virulence	Artificial stocking may increase virulence, and culling may reduce or increase virulence depending on pathogen life-history, culling selectivity and transmission dynamics	Frank (1996)
Seasonal drivers of disease emergence and dynamics	Intervention timing and frequency matters; control efforts can target transmission peaks	Altizer et al. (2006)
Pathogen interactions within hosts	Managing one pathogen alters the transmission and virulence of other pathogens	Fenton (2008)
Multi-host species disease dynamics	Reservoir hosts can drive the extinction of alternate hosts; rates of interspecific transmission may be inferred by managing one host species; management may need to target multiple host species	Dobson and Foufopoulos (2001)
Spread of disease in spatially structured hosts	Corridor vaccination can reduce disease in metapopulations; movement controls are unlikely to work for chronic infections	Keeling and Eames (2005)
Transmission increases with host density	Host density reductions may reduce disease transmission, and density thresholds for disease persistence may exist	Anderson and May (1979)
Transmission increases with disease prevalence independent of host density	Transmission associated with sexual interactions is more likely to cause host extinction, and non-selective culling may not reduce transmission	Getz and Pickering (1983)
Predation as a regulator of host population and disease	Predator conservation may reduce disease in prey populations	Packer et al. (2003)
Community composition, diversity and disease risk	Biodiversity loss and community disassembly may increase disease as predators and less-competent hosts are extirpated, depending on community composition and transmission dynamics	Keesing, Holt and Ostfeld (2006)
Environmental reservoirs and indirect transmission	Duration of disease control must scale with the environmental persistence; host extinction is more likely	Joh et al. (2009)
Individual-level variation and superspreading	Heterogeneity in individual resistance and infectiousness within a host population can lead to superspreaders that account for a large portion of transmission; management can target superspreaders	Lloyd-Smith et al. (2005)

Chapter 2

Does life history mediate changing disease risk when communities disassemble?

MAXWELL B. JOSEPH, JOSEPH R. MIHALJEVIC, SARAH A. ORLOFSKE, AND SARA H. PAULL

Biodiversity loss sometimes increases disease risk or parasite transmission in humans, wildlife and plants. Some have suggested that this pattern can emerge when host species that persist throughout community disassembly show high host competence—the ability to acquire and transmit infections. Here, we briefly assess the current empirical evidence for covariance between host competence and extirpation risk, and evaluate the consequences for disease dynamics in host communities undergoing disassembly. We find evidence for such covariance, but the mechanisms for and variability around this relationship have received limited consideration. This deficit could lead to spurious assumptions about how and why disease dynamics respond to community disassembly. Using a stochastic simulation model, we demonstrate that weak covariance between competence and extirpation risk may account for inconsistent effects of host diversity on disease risk that have been observed empirically. This model highlights the predictive utility of understanding the degree to which host competence relates to extirpation risk, and the need for a better understanding of the mechanisms underlying such relationships.

2.1 Introduction

Biodiversity loss alters ecosystem functioning and services (Cardinale et al., 2012). When local extinctions (extirpations) cause predictable losses of diversity, the consequences of local bio-

diversity loss are often non-random (Henle et al., 2004; Cadotte et al., 2011). In disease ecology, there is an emerging idea that host diversity can reduce or dilute disease risk, especially when diverse communities include species that reduce transmission due to low host competence—the ability to acquire and transmit infections (Keesing et al., 2010). Whether diversity generally decreases disease risk due to a dilution effect remains contentious partly due to uncertainty in whether host extirpation risk relates to host competence (Randolph and Dobson, 2012).

Covariance between host competence and extirpation risk could arise due to host life-history trade-offs, parasite adaptation to locally common hosts, or both. Host life-history trade-offs could operate if slow-lived species (e.g. those that are long-lived, large-bodied, and slowly reproducing) invest more heavily in parasite defence and experience higher extirpation risk, whereas fast-lived species invest little in parasite defence but are more ubiquitous in disassembled (or generally depauperate) communities (Huang et al., 2013b). Alternatively, parasites could adapt to locally common hosts, which likely experience lower extirpation risk due to higher population densities (Lively and Dybdahl, 2000). These mechanisms are not mutually exclusive, and both have received empirical support. However, the generality of these mechanisms remains uncertain, and the general consequences of resulting covariance between competence and extirpation risk have not been rigorously evaluated (but see Ostfeld and LoGiudice (2003) for a detailed system-specific evaluation).

To better understand how and why host diversity loss influences disease risk, we address three objectives: (1) clarify the rationale underlying extirpation risk-host competence relationships, (2) review the empirical evidence for these relationships and (3) evaluate the theoretical consequences for disease across a range of extirpation risk-host competence scenarios. Finally, we present empirical methods for disentangling the mechanisms underlying covariance between host competence and extirpation risk, with the goal of more accurately predicting the effects of biodiversity loss on disease dynamics in natural systems.

2.2 Determinants of Extirpation Risk

A variety of species traits such as reproductive pace, body size and home range size have been linked to global extinction risk, and are likely also important for local extirpation risk. Generally, large-bodied, slow-reproducing species recover more slowly from disturbances, and have low regional abundances which increases extirpation risk (Damuth, 1981; Hanski, 1991; Nee et al., 1991; Cardillo and Bromham, 2001; Meehan et al., 2004; Buckley and Jetz, 2007; Davidson et al., 2009). Furthermore, average home range size is negatively correlated with density in many mammals, and positively correlated to body size in birds (Bordes et al., 2009; Lee and Jetz, 2011), and as such, species with large home ranges and low population densities can experience higher extirpation risk (Boyle and Smith, 2010). However, among species with small populations, long-lived species appear to experience lower extirpation risk than those that are fast lived (Saether et al., 2011). Large body size, long lifespan, delayed maturity and specialised feeding behaviours can contribute to a higher frequency of local extirpation risk in fishes (Olden et al., 2007). Finally, species at the highest trophic levels are often more prone to local extirpations due to demography and direct human impacts such as hunting (Duffy, 2003). Such effects could lead to trophic cascades that influence not only species composition but also abundance in low diversity areas (Hebblewhite et al., 2005).

Ecological communities are often nested such that local species composition tends to be a nested-subset of the most species communities in a region (Wright, 1998). Variation in habitat characteristics (i.e. extrinsic factors) as well as species traits (i.e. intrinsic factors) has been implicated in the emergence of nested patterns (Almeida et al., 2007). The nested nature of many communities can also arise due to community disassembly at varying stages. For example, in Amazonian primate and carnivore communities, forest fragmentation leads to nested communities due to species-specific effects of patch size on extirpation risk (Michalski and Peres, 2005). To anticipate changes in host community competence as species are driven to local extinction, one could ask whether traits that mediate extirpation risk also mediate host competence.

2.3 Determinants of Host Competence

Host competence can be influenced by life-history trade-offs and parasite adaptation. Work within the field of ecoimmunology has suggested that relationships between host competence and life history traits could arise because defence against infection leads to an increase in the host's future reproduction (via enhanced survival) at the expense of current reproduction (Hawley and Altizer, 2010). In other words, investment in current and future reproduction trade-off due to limited resource allocation to growth, maintenance and reproduction (Stearns, 1992; Roff, 2001; Ricklefs and Wikelski, 2002). Therefore, individual investment in parasite defence is expected to co-vary with traits associated with survival and reproduction, such as body size, longevity and reproductive output (Lee, 2006; Cronin et al., 2010). This leads to the generalised prediction that fast-lived species invest less in immunological defences compared to slow-lived species (Figure 2.1; Cronin et al. (2010); Johnson (2012)). Such relationships between host competence and life history have been documented in the hosts of *Borrelia burgdorferi* (the causative agent of Lyme disease), West Nile Virus (Huang et al. 2013), trematode parasites (Johnson et al. 2012, 2013) and Kissing bugs (*Rhodnius pallescens*) that transmit *Trypanosoma cruzi* infections (Gottdenker et al., 2012). Host-parasite interactions (e.g. differential evasion of host immunity by parasites in a host community) could add variance to the relationship between investment in defence and realised host competence (Figure 2.1; Grenfell et al. (2004)). Nonetheless, it is reasonable to assume that increased investment in defence reduces host competence on average.

Parasite evolution provides another mechanism for the relationship between host traits and infection. Across host-pathogen systems, parasites exhibit local adaption to common host species and genotypes (Kaltz and Shykoff, 1998; Lively and Dybdahl, 2000; Lajeunesse and Forbes, 2002; Montarry et al., 2008). Likewise, due to the selective pressure of losing hosts during community disassembly, parasites that exploit multiple hosts may evolve to infect the most abundant or widespread hosts leading to the negative relationship between extirpation risk and host competence. Local adaptation of pathogens has received relatively limited attention in discussions of

diversity-disease relationships to date despite being suggested as a possible mechanism for disease dilution over a decade ago (Ostfeld and Keesing, 2000).

2.4 Covariance Between Extirpation Risk and Host Competence

Empirical evidence for competence-extirpation risk covariance is building. For three vector-borne disease systems, host species with fast life histories (e.g. small body size) that tend to experience lower extirpation risk are better able to acquire and transmit infections than slow-lived host species (Huang et al., 2013b). Similarly, fast-lived amphibians and those that are more abundant in species poor communities experience higher infection risk and more severe pathology from trematode parasites (Johnson, 2012; Johnson et al., 2013). This pattern appears relevant to zoonotic diseases as well: widely distributed primates with dense populations are more likely to transmit parasites to other species, including humans (Gómez et al., 2013). Still, departures from strict scaling between extirpation risk and host competence have been reported. A recent empirical study of primates did not detect statistically significant covariance between susceptibility to habitat disturbance and parasite prevalence a candidate proxy for host competence assuming constant parasite exposure (Young et al., 2013). Even when statistically significant relationships are detected, many papers report substantial variance in competence not explained by life history, which is to be expected due to imperfect covariance between investment in immunity and realised competence (Figure 2.1; Johnson (2012); Huang et al. (2013a)).

The general consequences of competence-extinction risk relationships have not been rigorously explored, despite being framed as an assumption and possible mechanism of the dilution effect (Randolph and Dobson, 2012; Huang et al., 2013a; Young et al., 2013). Beginning with an understanding from limited existing empirical data that there may be strong, weak or no correspondence between competence and extirpation risk, we built a model to explore how disease dynamics change as communities are disassembled under these possible scenarios. We predicted that strong negative covariance between extirpation risk and host competence would lead in most cases to dilution effects, because poor hosts are extirpated early in community disassembly, consistent with

previous studies that have made this assumption (Roche et al., 2012). As this covariance became weaker, we expected that the effects of species extirpations on disease risk would become more variable. Finally, we predicted that strong positive covariance (competent species experience higher extirpation risk) would tend to cause increased disease risk following extirpations (amplification effects) as community disassembly favors less-competent hosts.

2.5 Exploring the Consequences of Host Competence-Extirpation Risk Relationships

2.5.1 Model structure

We represent transitions from susceptible, infectious and recovered individuals in a multi-host community with a generalised directly transmitted microparasite using a system of differential equations (Dobson, 2004):

$$\frac{S_i}{dt} = (b_i - \delta N_i)N_i - d_i S_i - S_i \sum_{j=1}^N \beta_{ij} I_j \quad (2.1)$$

$$\frac{I_i}{dt} = S_i \sum_{j=1}^N \beta_{ij} I_j - I_i(d_i + v_i + \sigma_i) \quad (2.2)$$

$$\frac{R_i}{dt} = \sigma_i I_i - d_i R_i \quad (2.3)$$

where S_i , I_i and R_i represent the densities of susceptible, infectious and recovered individuals of species i . Recovered individuals are assumed to have life-long protective immunity. We assumed allometric scaling of host vital rates, disease-free equilibrium densities and infection induced death rates following established methods (Leo and Dobson, 1996; Dobson, 2004; Bolzoni et al., 2008). Details for each parameter are provided in Table 2.1.

Host competence was represented by the probability of transmission given a contact between a susceptible and infectious individual, which is proportional to the intraspecific transmission rate β_{ii} (Keesing et al., 2006). Intraspecific transmission rates were calculated based on the potential for an epidemic in a nave and isolated population (intraspecific R_0), recovery rates σ_i , pathogen-

induced death rates v_i , background mortality rates d_i and host densities at carrying capacity K_i (Table 2.1). Interspecific R_0 values were drawn from a right-skewed truncated gamma distribution so that populations of most species would not be readily invaded by the pathogen on their own, but some species would experience severe epidemics in the absence of other species, consistent with empirical evidence for multi-host pathogens (Komar et al., 2003; Ostfeld and LoGiudice, 2003; Huang et al., 2013a; Johnson et al., 2013). With strict allometric scaling, intraspecific transmission rates (β_{ii}) were linearly and positively correlated with intraspecific R_0 , and negatively related to body mass, so that small-bodied species were more competent than large-bodied species (Figure 2.4). Interspecific transmission between species j and species i (β_{ij}) was assumed to be symmetrical so that $\beta_{ij} = \beta_{ji}$, and was calculated as a scaled average of intraspecific transmission rates (Table 2.1; Dobson (2004)).

We used a two-step approach to explore the degree to which host competence-extirpation risk relationships affect changes in disease risk during community disassembly. First, we produced a gradient of communities spanning from the strict assumption of small-bodied species being most competent to the reverse assumption of small species being least competent. To achieve this, we first assumed reverse rank-ordering between body size and intraspecific R_0 , so that smaller bodied species received the highest values of R_0 (Roche et al., 2012, 2013). We then permuted the values of R_0 to produce all possible species- R_0 pairings, producing $N!$ unique communities all consisting of the same N species, but with different species- R_0 pairings. This created a gradient in the relationship between host competence and extirpation risk. On one extreme, smaller species always have the highest generated values of R_0 (i.e. a strong, negative correlation between host competence and body size). On the other extreme, smaller species always have lowest values of R_0 . Intermediate permutations represent a range of randomness (i.e. strength of covariance) from one extreme to the other, with maximum randomness occurring at the median number of inversions (Figure 2.5).

Second, we explored the consequences of competence-extirpation risk relationships in the context of community disassembly by sequentially removing individual host species from each of

the $N!$ communities generated in the previous step. We varied the extirpation rules to always, often or only rarely extirpate larger bodied species, gradually relaxing the empirically supported assumption that large-bodied (i.e. slow-paced) species have high extirpation risk. Specifically, we assumed (1) deterministic extirpations of largest-bodied hosts, (2) semi-random extirpations with extirpation risk proportional to $1w_i^{-1}$ and (3) completely random extirpations, irrespective of body size. The latter two methods represent partial and complete decoupling of extirpation risk and host competence. A range of initial host community richness values were explored, but our results were not sensitive to initial richness, and computational cost increased rapidly as host species were added. Thus, we present all model results with initial communities consisting of five species ($N = 5$).

We considered both compensatory and subtractive extirpations, corresponding to cases where persisting species do or do not increase in abundance following the extirpation of co-occurring species, respectively. With density compensation, following an extirpation, remaining species increased in density in proportion to their original relative abundances so that community density remained fixed and independent of species richness (MacArthur et al., 1972). Compensatory extirpations imply strong species interactions that result in host density regulation one mechanism for dilution (Keesing et al., 2006). Finally, in the interest of generality we considered both density-dependent transmission and frequency-dependent transmission, where host contact rates increase with or are independent of host densities, respectively. Density-dependent transmission is often seen with directly transmitted diseases, and frequency-dependent transmission is common with vector-transmitted diseases (McCallum et al., 2001). To capture general patterns emerging from the processes of randomly drawing intraspecific R₀ values from a distribution, and when applicable extirpation stochasticity, we iteratively produced and disassembled N! communities 1000 times for each extirpation scenario, assuming either density-dependent transmission or frequency-dependent transmission.

For each community, we estimated disease risk as community-level R₀, which is calculated as the dominant eigenvalue of the next-generation matrix (Dobson, 2004). This matrix incorporates

within-species and between-species transmission, as well as host species traits related to survival and reproduction. Community R_0 is positively related to the maximum prevalence of infection in the host community. Thus, using community R_0 as a measure of disease risk, we are able to incorporate both the ability of the pathogen to invade the host community, and the intensity of the resulting epidemic if invasion is possible. We evaluate the effect of community disassembly by quantifying changes in disease risk resulting from species extirpations: the difference in community R_0 before and after a species has been extirpated (ΔR_0). When a species is removed and ΔR_0 is positive, removing that particular species increases community-level disease risk; the opposite would be true if ΔR_0 is negative. Simulations were conducted with R version 3.0.1 (R Core Team 2013).

2.6 Results

As the assumption of perfect covariation between competence and extirpation risk was relaxed by permutation and/or stochastic extirpations, the effect of species extirpations on disease risk quickly became more variable, exhibiting both dilution- and amplification-effect behaviour in most cases. When large species were less competent and deterministically lost early in disassembly, dilution effects were observed in all cases except density-dependent transmission with subtractive extirpations (for which dilution effects are impossible). Even with deterministic extirpations, when the relationship between competence and life history was relaxed by permutation, both dilution and amplification effects could be observed for most cases as the variance in ΔR_0 increased rapidly (Figure 2.2). In general, large changes in disease risk following extirpations became more common across all cases when there was imperfect correspondence between life history and host competence (Figure 2.2).

With semi-random extirpations of large-bodied hosts, slight relaxation of the relationship between competence and life-history caused large increases in the overall variance in ΔR_0 so that both dilution and amplification effects were seen under most cases (Figure 2.2). However, as the relationship between competence and life history became more random, the increase in variance in

ΔR_0 was less pronounced for these semi-random extirpations than in the case of deterministic extirpations. With completely random extirpations, the overall variance in ΔR_0 was maximised, with a slight bias towards dilution, but with dilution and amplification nearly equiprobable regardless of the relationship between host competence and life history (Figure 2.6). When large-bodied species were more competent than small-bodied species and lost early in community disassembly, mixed amplification and dilution was still observed, and the expected effect of species losses on disease risk was sensitive to transmission dynamics and host-density compensation (Figure 2.2).

The mean effect of extirpation on disease risk was modulated by intensity of interspecific transmission (Figure 2.3). When c_{ij} is high, highly competent hosts cause greater increases in community-wide disease risk through increased rates of interspecific transmission. As the rate of interspecific transmission increased, the effects of species losses switched from dilution (on average) to amplification under all scenarios except density-dependent transmission with subtractive extirpations. This switch point occurred for lower rates of interspecific transmission as large-bodied species became more competent (Figure 2.3). When large hosts are more competent and extirpated early, the remaining host community is less competent on average. The degree to which large-bodied competent hosts increased risk for other host species depended on c_{ij} , causing an interaction between the among-species transmission rate and competence-life history relationship (Figure 2.3). In contrast, with density-dependent transmission and subtractive extirpations the magnitude of ΔR_0 increased with c_{ij} , but extirpations always reduced transmission potential due to a reduction in host community density.

Our findings were robust to the assumption of the existence of a recovered class with life-long protective immunity (analysis not shown), and to the distribution of host sizes in the community, which only slightly affected the direction of changes in disease risk following extirpation to an extent that was negligible relative to the effect of changing extirpation risk-competence relationships (Figure 2.7). Further, our results are qualitatively robust to the distribution of intraspecific R_0 in communities: with higher mean values of intraspecific R_0 , the mean values of ΔR_0 also increase, but the percentage changes in community disease risk are relatively consistent (Figure 2.8).

2.7 Discussion and future directions

Weak relationships between host competence and extirpation risk can cause mixed effects of species extirpations on disease risk. Further, the strength of the relationship, not simply the existence of a correlation, is a critical parameter for predicting how species extirpations affect the magnitude and direction of changes in disease risk. The predictions arising from this model extend earlier theoretical models predicting dilution effects with frequency-dependent transmission and amplification effects for density-dependent transmission with subtractive extirpations (Dobson, 2004; Rudolf and Antonovics, 2005; Ostfeld and Keesing, 2012). By explicitly incorporating variation and covariation in host competence and extirpation risk, this model suggests a novel and general explanation for why some empirical studies may fail to find evidence for the dilution effect in scenarios where it previously would have been predicted to occur (Randolph and Dobson, 2012; Wood and Lafferty, 2012; Salkeld et al., 2013).

With strong negative covariance between host competence and extirpation risk, we would predict that generally, dilution effects are much more likely in all cases except that of purely subtractive extirpations and density-dependent transmission. In nature, communities with subtractive extirpations would hypothetically be composed of non-interactive species that utilise very different resources and share no natural enemies, which would likely lead to very low contact rates that severely limit interspecific transmission of a density-dependent parasite. Thus, truly generalist pathogens with density-dependent transmission in real communities with subtractive extirpations may be rare. Further, compensatory extirpations imply very strong species interactions that could be associated with high interspecific contact rates under density-dependent transmission. Depending on the relationship between host competence and host life history, such highly interactive communities could experience net dilution or amplification that empirically appears idiosyncratic due to high variability. Taken together, subtractive and compensatory extirpations represent two extremes, and considering more realistic intermediate cases is an important frontier in diversity-disease research. Similarly, while we considered the consequences of strong positive covariance

between extirpation risk and host competence, this kind of relationship remains much less empirically supported than the reverse case, where fast-lived species are more competent and less prone to extirpation.

While this model predicts the consequences of competence-extirpation risk relationships, empirical studies are necessary to address the mechanisms underlying competence-extirpation risk covariance. For instance, experimental infections of geographically and genetically distinct host populations with either sympatric or allopatric parasites can reveal the extent to which parasite adaptation to locally common hosts has occurred (Lively and Dybdahl, 2000). Similarly, to test whether host life-history trade-offs influence host competence in a community, studies could evaluate how strongly resource allocation to parasite defence trades-off with allocation to reproduction and other maintenance expenditures (Martin et al., 2006). Reciprocal pathogen transplants crossed with manipulations of host investment in life-history components (e.g. clutch augmentation) that trade-off with parasite defence could reveal interactions between these two mechanisms. Lastly, community disassembly imposes a changing fitness landscape for generalist pathogens as hosts disappear or vary in abundance. The selective pressures on symbiont communities imposed by host community disassembly remains an open area for future research, with the potential to further understanding of the evolution of host specificity and virulence.

Mechanisms that decouple host competence from extirpation risk may be as important as those that cause them to covary. These include ecological processes such as stochastic extirpations, overexploitation and trophic interactions, as well as methodological issues such as ignored covariate uncertainty. Stochasticity in community disassembly could reduce the strength of competence-extirpation relationships, and can result from local extrinsic factors such as extreme weather events that disproportionately affect species that would otherwise not be at the greatest risk of extirpation (Lande, 1993). In addition, the particular order of extirpations among small populations may be difficult to predict because of demographic stochasticity, suggesting that competence-extirpation risk relationships may be more detectable in host communities with a wide range of population sizes. Overexploitation can further decouple extirpation risk from life history considering that even

abundant species can be extirpated given enough hunting pressure (Gaston and Fuller, 2008). Further, loss of predators early in community disassembly may have strong effects on prey density and pathogen transmission via trophic cascades and predator release (Ostfeld and Holt, 2004; Dobson et al., 2006; Bruno and Cardinale, 2008; Levi et al., 2012). Methodologically, host competence and extirpation risk are typically estimated with error. If the error in both quantities is not accounted for, estimates of the effects of extirpation risk on host competence (or vice versa) will be biased towards zero due to an error-in-covariate problem (Pearson, 1901). When covariance between extirpation risk and host competence is not detected even after controlling for uncertainty in both, it is important that results are published and available for future meta-analyses.

2.8 Conclusions

Deviations from strict scaling between extirpation risk and host competence introduce a large amount of variability in the effect of species extirpations on community disease risk, and strict negative scaling between extirpation risk and competence leads to dilution effects in most cases. Such deviations may explain idiosyncratic changes in disease risk. Whether host competence generally corresponds to extirpation risk remains an active research topic, but we show that the existence, and more importantly the strength, of such correspondence can strongly mediate the magnitude, direction and variability in the effect of species extirpations on disease risk. To understand the mechanistic basis for competence-extirpation risk relationships, it will likely be necessary to broaden empirical work to account for life-history trade-offs and pathogen adaptation. In general, efforts to directly link extirpation risk with host competence, rather than relying on proxy measures such as host life history, will help to clarify the level of variability that can be expected in extirpation-competence relationships, and thus the predicted generality of the dilution effect. Finally, this study highlights the value of linking variation in host competence to extinction risk at multiple scales to better understand local and global effects of biodiversity loss on ecosystem services.

Table 2.1: Parameter definitions

Parameter	Definition	Value
w_i	Body mass of species i (kg)	aw_{min}^{i-1}
w_{min}	Minimum species body mass	1 kg
a	Weight scaling parameter	varied from 1 to 5
b_i	Per capita birth rate	$0.6w_i^{-0.27} + 0.4w_i^{-0.26}$
d_i	Per capita death rate	$0.4w_i^{-0.26}$
K_i	Carrying capacity	$(b_i - d_i)/\delta$
δ	Density dependence intensity parameter	0.01
v_i	Infection-induced mortality rate	$(m - 1)d_i$
m	Virulence term	1.5
σ_i	Recovery rate	rd_i
R_{0i}	Intraspecific pathogen reproductive ratio	$\text{Gamma}(k, \theta) \in (0, 10)$
k	Gamma shape parameter	varied from 0.1 to 1
θ	Gamma scale parameter	varied from 1 to 2
β_{ii}	Intraspecific transmission rate	$K_i^{-1}R_{0i}(d_i + v_i + \sigma_i)$
β_{ij}	Interspecific transmission rate	$c_{ij} \frac{\beta_{ii}\beta_{ij}}{2}$
c_{ij}	Interspecific transmission scaling parameter	varied from 0.01 to 0.99

2.9 Acknowledgements

We thank the Johnson Laboratory at the University of Colorado-Boulder and three anonymous reviewers for providing invaluable comments on this manuscript. MBJ and JRM were supported by the National Science Foundation Graduate Research Fellowship Programme.

Figure 2.1: Life history can drive covariance between competence and extirpation risk via the component relationships illustrated in panels AC, but this covariance is less likely to emerge as the variability around the component relationships increases. Data for a 10 species community are simulated by drawing life-history pace values from a normal distribution. These values are used to generate species-specific extirpation risk and investment in parasite defence according to a negative linear model, with the open red points having more variability than the closed black points. Investment in defence is similarly used to generate species-specific values of host competence based on a simple linear model. Subsequent relationships between extirpation risk and host competence are plotted in panel D. All panels show least-squares linear regression lines with 95

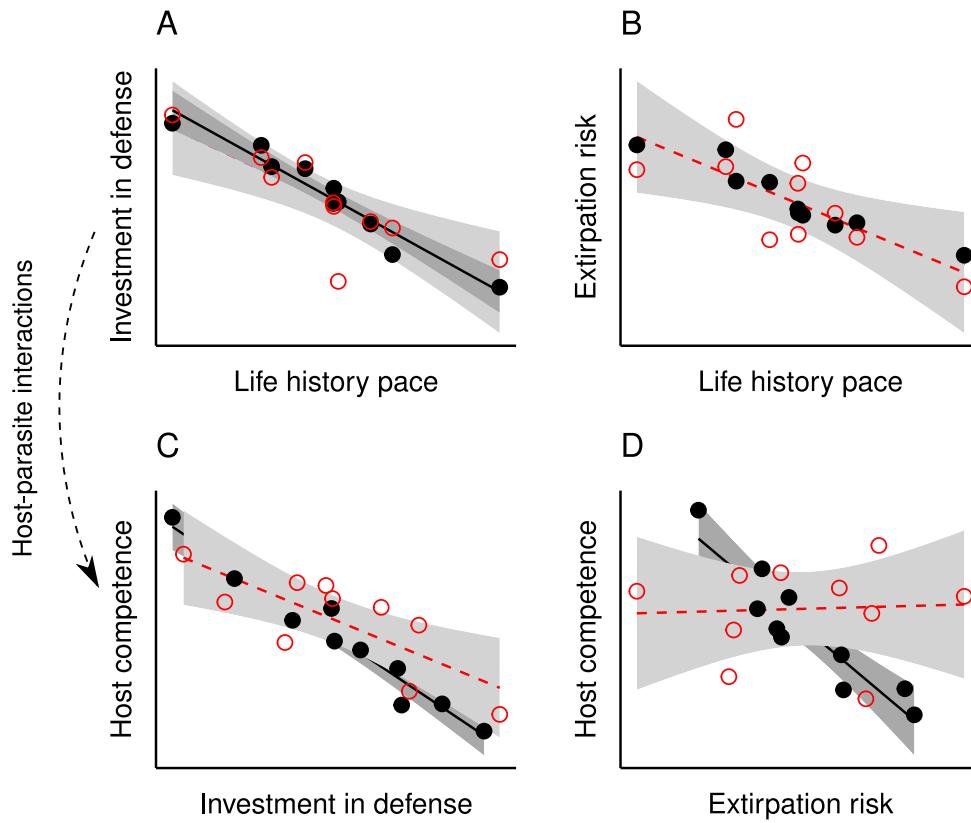


Figure 2.2: Split violin plots show the distributions of the effects of extirpations on community R_0 across 1000 iterations with deterministic extirpations of large-bodied species in white, and stochastic extirpations with risk proportional to $(1w_i^1)$ in grey. Red represents amplification (higher community R_0 in the more diverse community); blue dilution (lower community R_0 in the more diverse community). The leftmost distributions correspond to the identity permutation of intraspecific R_0 values (perfect negative rank scaling between extirpation risk and body size) and distributions to the right contain progressively more inversions, until perfect positive rank scaling between host competence and body size is achieved in the rightmost distributions. Intermediate distributions represent a gradient between these two extremes. Horizontal black lines indicate median effects, with surrounding bars showing 25 and 75

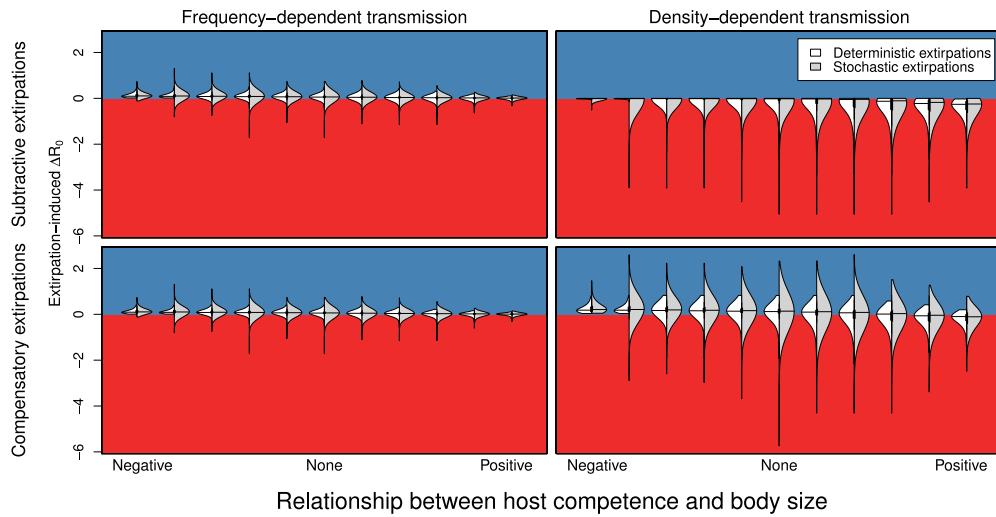


Figure 2.3: Contour plot illustrating variation in the mean effect of species extirpations on community R_0 across interspecific transmission strengths and competence-extirpation risk relationships. Larger values of c_{ij} lead to greater interspecific transmission rates (Table 2.1). Here, large-bodied species are deterministically extirpated, coupling body size with extirpation risk. The x-axis is identical to that of Fig. 2. Numbers along contours denote mean ΔR_0 isoclines, with the zero-isocline marked by a dashed line. Red represents amplification; blue dilution. Here, $a = 3$, $k = 0.5$, $\theta = 1.5$, and c_{ij} was varied from 0.01 to 0.99.

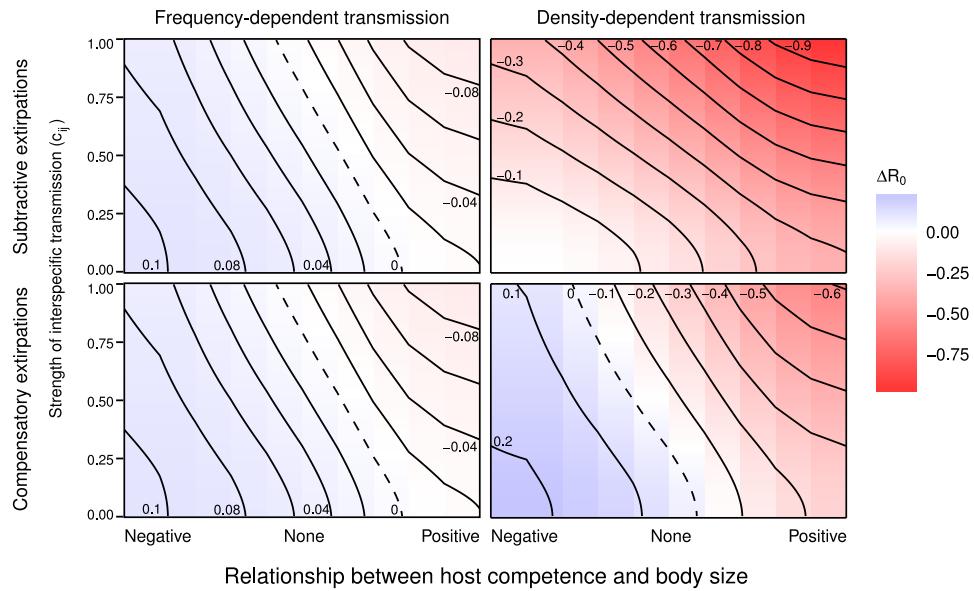


Figure 2.4: Line graph depicting relationships between transmission and weight across iterations employing strict scaling (smaller bodied organisms always transmit at a higher rate).

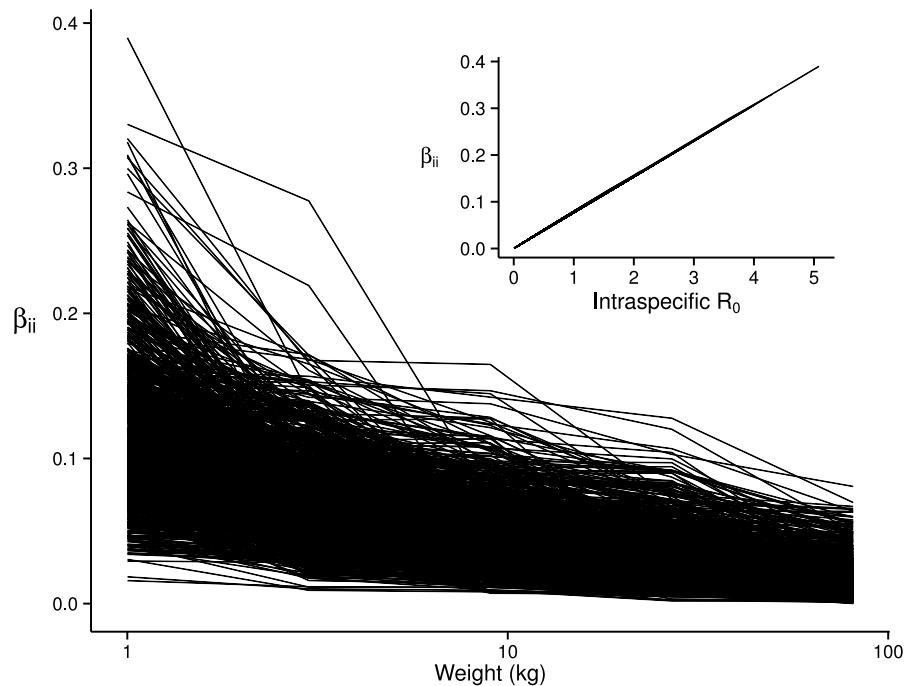


Figure 2.5: Scatter plots with overlaid linear regression lines depicting the range of relationships between transmission and weight across iterations as a function of the number of inversions in transmission-weight tuples.

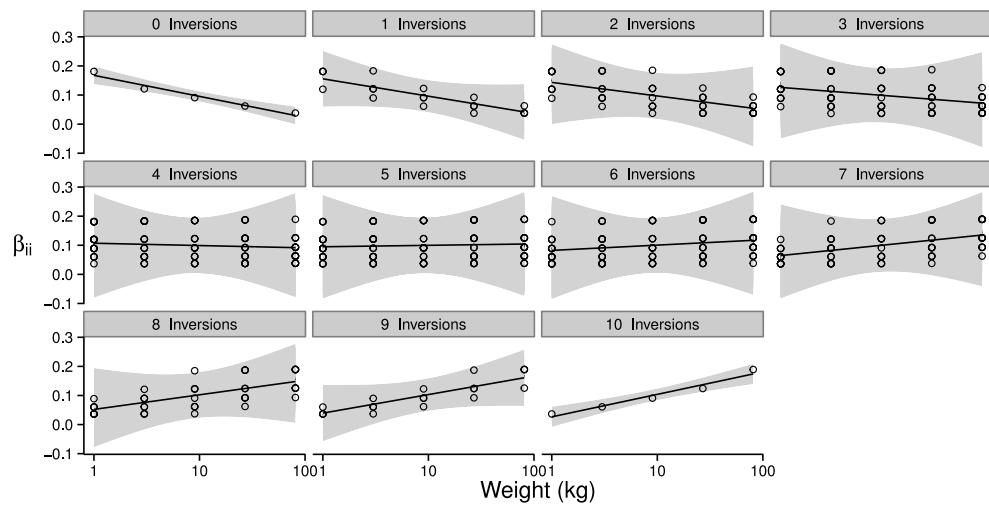


Figure 2.6: Split violin plots show the distributions of the effects of extirpations on community R_0 across 1000 iterations with deterministic extirpations of large-bodied species in white, and totally random extirpations in grey. Red represents amplification (higher community R_0 in the more diverse community); blue dilution (lower community R_0 in the more diverse community). The leftmost distributions correspond to the identity permutation of intraspecific R_0 values (perfect negative rank scaling between extirpation risk and body size) and distributions to the right contain progressively more inversions, until perfect positive rank scaling between host competence and body size is achieved in the rightmost distributions. Intermediate distributions represent a gradient between these two extremes. Horizontal black lines indicate median effects, with surrounding bars showing 25 and 75

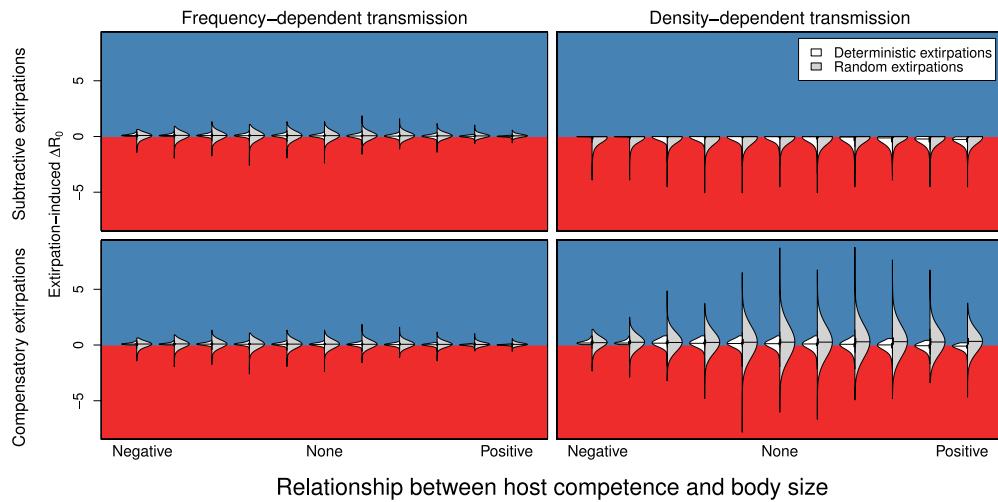


Figure 2.7: Contour plot showing the sensitivity of model predictions to the body size scaling parameter. Solid lines represent ΔR_0 isoclines as in 2.3.

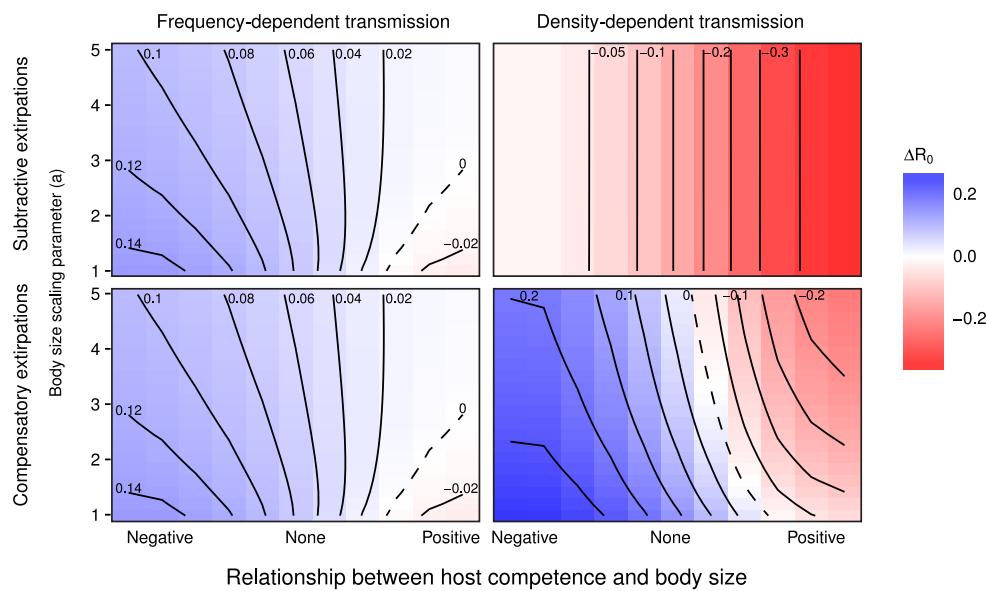
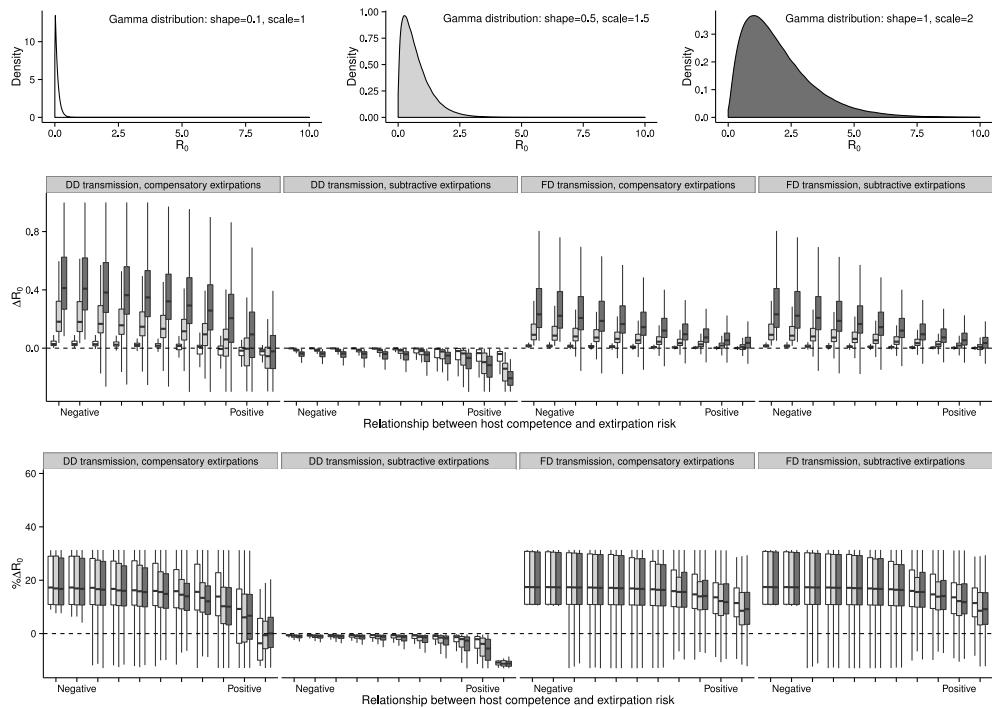


Figure 2.8: Model sensitivity to the distribution of R_0 values among species. We chose from among three different truncated gamma distributions with varying means and variance.



Chapter 3

Integrating occupancy models and structural equation models to understand species occurrence

MAXWELL B. JOSEPH, DANIEL L. PRESTON, PIETER T. J. JOHNSON

Understanding the drivers of species occurrence is a fundamental goal in basic and applied ecology. Occupancy models have emerged as a popular approach for inferring species occurrence because they account for problems associated with imperfect detection in field surveys. Current models, however, are limited because they assume covariates are independent (i.e., indirect effects do not occur). Here, we combined structural equation and occupancy models to investigate complex influences on species occurrence while accounting for imperfect detection. These two methods are inherently compatible because they both provide means to make inference on latent or unobserved quantities based on observed data. Our models evaluated the direct and indirect roles of cattle grazing, water chemistry, vegetation, nonnative fishes, and pond permanence on the occurrence of six pond-breeding amphibians, two of which are threatened: the California tiger salamander (*Ambystoma californiense*), and the California red-legged frog (*Rana draytonii*). While cattle had strong effects on pond vegetation and water chemistry, their overall effects on amphibian occurrence were small compared to the consistently negative effects of nonnative fish. Fish strongly reduced occurrence probabilities for four of five native amphibians, including both species of conservation concern. These results could help to identify drivers of amphibian declines and to prioritize strategies for amphibian conservation. More generally, this approach facilitates a more mechanistic representation of ideas about the causes of species distributions in space and time.

As shown here, occupancy modeling and structural equation modeling are readily combined, and bring rich sets of techniques that may provide unique theoretical and applied insights into basic ecological questions.

3.1 Introduction

Much of ecology is concerned with explaining and predicting where species occur in space and time. While species distributions are directly and indirectly influenced by a diverse suite of abiotic and biotic factors, most statistical treatments do not formally differentiate between direct and indirect effects (Guisan and Thuiller, 2005). Structural equation modeling (SEM) provides one means by which to formally represent causal assumptions in the form of direct and indirect causal pathways, helping to close the gap between biological mechanisms and statistical methodology (Bollen, 1989; Grace, 2006).

Structural equation models combine causal assumptions with observations to generate causal inferences (Pearl, 2000a). Often, these assumptions are displayed in a path diagram that represents model components and causal relationships (i.e. how one variable affects another). These path diagrams encode both direct and indirect effects, such as those involved in complex processes that structure communities (Clough, 2012; Alsterberg et al., 2013). As a result, SEM provides a framework for clarifying, representing, and evaluating hypotheses in ecology that can outperform traditional associational statistical methods. For example, the relationship between productivity and biodiversity has historically been evaluated statistically in a bivariate context, despite the multidimensionality of the hypotheses that have been put forward to explain the relationship. By explicitly representing causal relationships between productivity and diversity, along with disturbance and stressors in a SEM, the mechanistic relationship between productivity and biodiversity becomes more clear, with more predictive power than bivariate approaches (Grace et al., 2014). In addition, SEM is flexible, with recent developments facilitating non-linear effects and a wide range of distributions, broadening the potential applicability of SEM in ecology (Lee, 2007). Finally, SEM differentiates between latent and observable variables. Latent variables cannot be observed

directly, but observations can provide information on latent quantities. A classic example is the presence or absence of a species.

In practice, studies of species occurrence are notoriously plagued by imperfect detection because species may be present but unobserved. This can be problematic when trying to explain species distributions because false absences create bias in estimated coefficients for covariates that explain occurrence. For instance, if turbidity increases the probability of occurrence of an aquatic organism, but decreases detectability (i.e. the organism is harder to observe), detection events will not appear to be related to turbidity. However, adjusting for imperfect detection facilitates unbiased estimation of the true relationship between such a factor and species occurrence. Such adjustments can be made via occupancy modeling, a method in which repeat surveys are used to estimate probabilities of detection and draw inferences on true occupancy states (MacKenzie et al., 2002). Occupancy models are classic hierarchical models, where the latent binary occurrence state is partly observed, and the extra zeros in detection histories are modeled as potentially arising from a combination of true and false absences. As the appreciation and application of occupancy models has increased, they have been extended to include multiple species, temporal dynamics including colonization and extinction, estimation of species richness, habitat ephemerality, and even disease dynamics (Dorazio et al., 2010; MacKenzie et al., 2011; Miller, 2012). These developments have been instrumental in improving inferences about the processes that govern species distributions. However, many existing models require independent covariates for occurrence and detection. This limitation represents a challenge to understanding even moderately complex processes that drive species distributions that may include a combination of direct and indirect effects on occurrence and detectability.

Here, we show that combining SEM with occupancy models overcomes this limitation and provides a means by which to more directly represent processes that influence species occurrence. We aim to illustrate the practical and conceptual compatibility of occupancy modeling and structural equation modeling by developing a causal model of species occurrence and community composition that explicitly accounts for non-detection. We use pond-breeding amphibians as an applied case

example. First, we outline the logic behind the model by summarizing the relevant empirical knowledge for our system in terms of the effects of pond permanence, cattle grazing, nonnative fishes, water chemistry, and vegetation on amphibian occurrence. We then discuss the data collection scheme and formally relate our observations to the processes under consideration by developing observation, process, and parameter (prior) models. As a model verification step, we demonstrate the recovery of parameters from our SEM using simulation (Rykiel, 1996). Last, we present results for the case study and discuss the value, limitations, and future directions for this approach.

3.2 Study system

The amphibian community in freshwater ponds of the San Francisco Bay Area of California, USA includes the Pacific chorus frog (*Pseudacris regilla*), western toad (*Anaxyrus boreas*), California newt (*Taricha torosa*), California tiger salamander (*Ambystoma californiense*), California red-legged frog (*Rana draytonii*), and nonnative American bullfrog (*Lithobates catesbeianus*) (Stebbins, 2003). Of the native species, two are federally protected and are of broad conservation concern (*A. californiense* and *R. draytonii*), and *A. boreas* has declined in specific regions (Lannoo, 2005). Some populations of these species are currently threatened by invasive species including nonnative fishes and bullfrogs, which potentially act as predators and competitors for native amphibians (Fisher and Shaffer, 1996; Lawler et al., 1999; Preston et al., 2012). Further, widespread cattle grazing in this region has the potential to affect the population dynamics of amphibians at breeding sites directly via trampling, and indirectly by removing vegetation and increasing the concentration of nitrogenous compounds in wetlands (Robins and Vollmar, 2002; Roche et al., 2012). The ponds we considered included both temporary and permanent water bodies up to 7 meters in depth.

3.3 Data collection

From 2009 to 2013, field crews used standardized methods to assess amphibian site occupancy in 171 wetlands in the San Francisco Bay Area, USA (Contra Costa, Alameda, Santa Clara, and

San Mateo counties) (Johnson et al., 2013). For simplicity, we used only one year of data from each site (with years chosen at random). The sites consisted of both artificially created livestock ponds and natural wetlands. Crews surveyed wetlands twice per summer using a combination of visual encounter, seine, and dipnet surveys (Crump and Bury, 1994). Visual encounter surveys involved a single observer walking the perimeter of the wetland and recording all species observed. Dipnet sweeps were conducted every 3 m around the pond perimeter (1.4 mm mesh and 2600 cm² net area) and three to four seine net hauls were performed in the deepest regions of each pond (4 mm mesh size and 1 x 2 m net area). For each survey, we recorded all life stages of each observed amphibian species. We considered sites “occupied” if larval amphibians were present, which indicates adult presence, successful fertilization, some degree of larval development, and potential for metamorphosis (i.e. breeding activity).

We quantified cattle grazing intensity, percentage of shoreline vegetated, pond permanence, water chemistry, and the presence of nonnative fish. To measure cattle grazing intensity, we recorded the number of cow paddies within three meters of shoreline and recorded shoreline perimeter with a handheld GPS unit to quantify cow paddy density at each pond. As a second measure of grazing intensity, we made a qualitative judgment of whether the wetland was disturbed by cattle based on evidence of tracks and trampling. Water samples were collected in acid-washed Nalgene bottles, filtered, frozen, and analyzed for ammonium (NH4⁺) and total dissolved nitrogen (TDN) concentrations using standard protocols (see: <http://snobear.colorado.edu/Kiowa/Kiowaref/procedure.html>). Imagery from Google Earth was used to help determine whether wetlands were permanent, based on consistent presence of water year round. Last, crews estimated the percentage of pond shoreline that was vegetated and recorded whether fish were present or absent using observations from all sampling methods. The most common fish species were nonnative, including mosquitofish (*Gambusia affinis*), largemouth bass (*Micropterus salmoides*), and bluegill sunfish (*Lepomis macrochirus*).

3.4 Conceptual model

Drawing upon previous literature, we developed a multivariate hypothesis about the drivers of amphibian community composition in this system (Figure 3.1). We expected fish to strongly affect community composition, particularly species with poor avoidance strategies and high palatability (Kruse and Stone, 1984; Kats et al., 1988; Adams, 2000; Welsh et al., 2006). Further, pond permanence should increase bullfrog occurrence because bullfrogs have a multi-year larval development period that is longer than most western native species (Collins, 1979). We hypothesized that livestock would physically alter wetland ecosystems via trampling and grazing, and chemically alter wetlands via inputs of nitrogenous waste products in urine and feces (Adams et al., 2009; Schmutzler et al., 2008; Jansen and Healey, 2003; Knutson et al., 2004; Kauffman et al., 1983). Previous studies indicate that characteristics of shoreline vegetation influence amphibian breeding and that water chemistry can alter reproductive success and occupancy probability (Earl and Whitteman, 2009; Burne and Griffin, 2005; Egan and Paton, 2004; Jansen and Healey, 2003; Brodman et al., 2003; Rouse et al., 1999; Rowe and Dunson, 1995; Freda and Dunson, 1986). Separating out the causal pathways through which grazing alters amphibian occupancy is important from a practical standpoint, as each one might be targeted differently with management interventions (e.g., limiting cattle-induced damage to vegetation versus limiting the total intensity of grazing).

3.5 Model formalization

To formalize the model of Figure 3.1, we adopted a hierarchical Bayesian approach, developing model components related to observations, processes in the system, and priors for unknown parameters (Lee, 2007; Cressie et al., 2009; Dorazio et al., 2010).

3.5.0.1 Observation model

While early structural equation models assumed normally distributed observed indicator variables, current methods are extremely flexible in terms of likelihood functions and relationships

between variables, broadening the applicability of these methods in ecology (Grace, 2006; Lee, 2007). We consider the presence/absence state of a species at a location to be a hidden (latent) binary variable which gives rise to imperfect detection/nondetection data. The use of a latent state variable is a key conceptual link between the fields of SEM and occupancy modeling. The flexibility of this approach is demonstrated in the following sections where we apply a variety of observation models from the exponential family, a moderately complex finite-mixture distribution, and a combination of continuous and discrete latent variables.

We considered grazing intensity by cattle to be a continuous latent quantity that cannot be observed directly, whose value is indicated by evidence of disturbance and the density of cow paddies in the vicinity of a wetland. Specifically, cattle disturbance was modeled as a Bernoulli random variable with latent grazing intensity (ξ) as a continuous covariate:

$$Y_1[j, k] \sim \text{Bernoulli}(p_{Y_1}[j])$$

$$\text{logit}(p_{Y_1}[j]) = \beta_{Y_1,0} + \beta_{Y_1,1}\xi[j]$$

for the j^{th} site and the k^{th} survey, with square brackets representing indexing. Observed cow paddy counts are treated as a second indicator (specifically, a multi-method indicator) (Grace, 2006) and modeled as a Poisson random variable with an offset for shoreline perimeter of site j ($\mu_{perim}[j]$) in meters:

$$Y_2[j, k] \sim \text{Poisson}(\lambda[j, k])$$

$$\frac{\lambda_{j,k}}{\mu_{perim}[j, k]} = e^{\beta_{Y_2,0} + \beta_{Y_2,1}\xi[j]}$$

Perimeter observations are subject to measurement error and variation within a season. Therefore, we modeled latent mean perimeter values for each site (μ_{perim}), which represents the expected pond perimeter value for site j :

$$\log(\mu_{perim}[j, k]) \sim N(\mu_{perim}[j], \sigma_w)$$

$$\mu_{perim}[j] \sim N(\alpha_{perim}, \sigma_a)$$

where σ_w represents measurement error and variation within a season, α_{perim} is the (log) average perimeter across all wetlands, and σ_a is the variability in perimeter among wetlands.

Shoreline vegetation, v , ranged from 0-100% with a non-negligible number of 0%, 100%, and intermediate observations. Therefore, we treated these observations as arising from a zero-one inflated beta distribution, which is a finite mixture distribution with a Bernoulli component that produces 0's and 1's, and a beta component that produces values on the interval (0, 1) (Ospina and Ferrari, 2012):

$$P(v; \alpha, \mu_v, \phi) = \begin{cases} \alpha(1 - \mu_v) & v = 0 \\ \alpha\mu_v & v = 1 \\ (1 - \alpha)f(v; \mu_v, \phi) & 0 < v < 1 \end{cases}$$

$$logit(\mu_v[j]) = \beta_{v,0} + \beta_{v,1}\eta_1[j]$$

$$logit(\alpha) = a_0 + a_2\mu_v^2$$

Here, α determines the extent to which the beta or binomial mixture components dominate the probability density function. The second degree polynomial term with coefficient a_2 causes extreme values of the logit-expected shoreline vegetation cover μ_v to increase the probability of an observer recording either 0% or 100% shoreline vegetation cover. This formulation essentially imposes a minimum probability of observers recording a discrete value when the true shoreline vegetation is 50%, and increases the probability of 0% or 100% observations as the true cover approaches those values. Last, $f(v; \mu_v, \phi)$ is the probability density function of the beta distribution, parameterized in terms of its mean (μ_v) and variance (ϕ), which represents the combination of observation error and within-summer variation in true shoreline vegetation cover (Ospina and Ferrari, 2012).

Log-transformed concentrations of ammonium and total dissolved N in the water were used as multi-method indicators of latent N concentration. Because both total dissolved N and NH_4^+ concentrations tend to increase over the course of the summer due to exogenous inputs and pond drying, we also included an effect of survey number k , coded as an indicator variable for the 2nd survey, such that $I(k = 2) = 1$, and $I(k = 1) = 0$.

$$\log(NH_4^+[j, k]) \sim N(\beta_{NH_4^+, 0} + \beta_{NH_4^+, 1}\eta_2[j] + \beta_{NH_4^+, 2}I(k), \sigma_{NH_4^+})$$

$$\log(N[j, k]) \sim N(\beta_{N, 0} + \beta_{N, 1}\eta_2[j] + \beta_{N, 2}I(k), \sigma_N)$$

We adopt an occupancy modeling approach for the data model describing amphibian detection and non-detection. Observations of the i^{th} species at the j^{th} site on the k^{th} repeat survey are represented by $Y[i, j, k]$. We treated these observations as Bernoulli random variables with probability $p[i, j, k]z[i, j]$, where p is the probability of detection and z is the latent binary presence/absence state. The true occurrence state z is only partly observed. If species i was seen at site j on any survey, it was present, but if it was not seen on any survey, it is possible that it was present but unobserved (MacKenzie et al., 2002):

$$Y[i, j, k] \sim Bernoulli(p[i, j, k]z[i, j])$$

We treated fish presence and pond permanence as directly observed quantities, which is supported by consistency of observations within and across years in this system.

3.5.0.2 Process model

Our process model represents the latent processes connecting the latent quantities: grazing intensity ξ , shoreline vegetation η_1 , N concentration η_2 , the true occupancy states z , and the probability of detection P (Figure 3.2). In the SEM lexicon, process models are sometimes referred to as structural models. Note that there is an additional pathway from fish to nitrogen that was

not initially hypothesized, but revealed in the process of model evaluation (see Model Assessment below).

We treat grazing intensity as an exogenous latent variable, unaffected by the other latent quantities, with mean 0 and standard deviation 1 as an identifiability constraint. Due to cattle grazing on and trampling of shoreline vegetation, we modeled a linear effect of cattle grazing intensity ξ on shoreline vegetation η_1 :

$$\eta_1 \sim N(\gamma_1 \xi, 1)$$

where the standard deviation term, similarly set to 1 for identifiability, represents the influence of other, unmodeled factors on shoreline vegetation cover. Nitrogenous inputs from cattle excretion in and around wetlands are treated similarly. Following a graphical check of independence assertions, we included an effect of fish presence (γ_3) on nitrogen:

$$\eta_2 \sim N(\gamma_2 \xi + \gamma_3 fish, 1).$$

We represented true occupancy states as Bernoulli random variables with probability of occupancy $\psi[i, j]$ for the i^{th} species at the j^{th} site, such that $z[i, j] \sim Bernoulli(\psi[i, j])$. We use a logit-link to model the effects of observed and latent covariates on ψ :

$$logit(\psi[i, j]) = \alpha_{species}[i] + \alpha_{region}[i, r[j]] + \alpha_{site}[i, j]$$

where $\alpha_{species}$ accounts for species-specific differences in overall occupancy across the entire study area, α_{region} accounts for regional differences in occupancy rates within each species (with region r being indexed by site j). Last, α_{site} represents the local effects of fish, pond permanence, grazing, shoreline vegetation and nitrogen concentrations. These terms can be decomposed as follows:

$$\alpha_{region}[i, j] \sim N(0, \sigma_{\alpha_{region}}[i])$$

This varying intercept term accounts for among-region variation within species in occupancy, with varying species-specific standard deviations, to account for the fact that the degree of regional variation in occupancy rates varies among species. Local covariate effects enter the final term:

$$\alpha_{site}[i, j] = \beta_\psi[i, 1]fish[j] + \beta_\psi[i, 2]perm[j] + \beta_\psi[i, 3]\eta_1[j] + \beta_\psi[i, 4]\eta_2[j] + \beta_\psi[i, 5]\xi[j]$$

Species vary in their detection probabilities and p may also vary between first and second visits:

$$\text{logit}(p[i, j, k]) = \alpha_p[i] + \beta_p k$$

where α_p is a species-specific mean and the last term represents the effect of early vs. late summer surveys.

3.5.0.3 Priors

We assumed logit-normal species responses to covariates with covariate-specific community means and variance parameters that represent among-species variability, such that:

$$\beta_\psi \sim N(\mu_{\beta_\psi}, \sigma_{\beta_\psi})$$

$$\beta_p \sim N(\mu_{\beta_p}, \sigma_{\beta_p})$$

Further, we assumed that for each species, mean detection probabilities would be logit-normally distributed around community level means:

$$\alpha_p \sim N(\mu_{\alpha_p}, \sigma_{\alpha_p})$$

Hierarchical parameters corresponding to community-level variance terms received semi-informative half-Cauchy priors that were weighted towards small values to reduce bias relative to commonly used uniform priors (Gelman and Pardoe, 2006). We adopt vague priors for all other

parameters except the loading terms for indicator variables which were constrained to be positive (e.g., increases in latent grazing intensity ξ correspond to increases in its indicators Y_1 and Y_2). Last, prior information based on previous work has entered the model in the form of included effect pathways (e.g., we assume that amphibian community composition does not affect grazing intensity).

3.6 Estimation

We used Stan, and the R package rstan to draw samples from the joint posterior distribution of all parameters (Stan Development Team, 2014c; R Core Team, 2014; Stan Development Team, 2014a,b). Although this model could be implemented with WinBUGS, OpenBUGS, or JAGS, the simultaneous updating of all parameters via the No-U-Turn sampler in Stan results in faster convergence and more efficient sampling (Hoffman and Gelman, 2014). Running three chains in parallel with 30000 iterations took about 80 minutes on a quad-core i7 laptop. Convergence was assessed using visual inspections of trace plots and the Gelman-Rubin potential scale reduction factor (Gelman and Rubin, 1992; Brooks and Gelman, 1998).

3.7 Model assessment

As this is a new method, we conducted a simulation analysis to ensure adequate recovery of parameter estimates across a range of known values. We simulated approximately 100 datasets with structure identical to our model, and the same amount of information (observations) present in our dataset. We then attempted to recover the known parameters by fitting the model to our simulated datasets (Gimenez et al., 2012). Any simulations that did not reach convergence at the MCMC step were discarded.

Independence assumptions were evaluated graphically using scatter plots to detect missing causal pathways (Grace et al., 2012). This revealed a positive correlation between fish presence and N (both total dissolved N and NH_4^+), leading to the inclusion of an additional effect of fish on N that we had not initially hypothesized. Mechanistically, this pathway may represent the joint

effects of fish locking up N in their tissues, and suppressing zooplankton through predation, leading to release of phytoplankton from grazing and greater uptake of N from the water (Andersson et al., 1978; Henrikson et al., 1980).

3.8 Results

3.8.1 Parameter recovery

Our simulations demonstrated parameter recovery with 95% highest density posterior intervals (HDIs) including the true population level parameters over 90% of the time (Figure 3.5). Furthermore, we were able to recover effects of local factors (β_ψ), identifying large effects as being non-zero (Figure 3.6). These results increased confidence in parameter estimates for the empirical data and gave an indication of the power that we might have to detect effects of local drivers of occurrence with a dataset of comparable size.

3.8.2 Empirical results

Nonnative fish and pond permanence directly affected amphibian community composition, with fish exerting the most consistent and strongest effects. Nonnative fish reduced the probability of occurrence for four of five native amphibians (all species except western toads and bullfrogs; Figure 3.3). Species-specific effects were observed for pond permanence, which had a strong positive effect on bullfrog and red-legged frog occurrence. Shoreline vegetation, nitrogenous compounds, and cattle grazing exerted relatively weak direct effects, with all HDIs including zero. Similarly, all HDIs for indirect effects of fish (via nitrogen) and cattle (via nitrogen and shoreline vegetation) on amphibian occurrence included zero.

Consistent with our expectations and previous work, cattle grazing decreased shoreline vegetation (HDI: (-0.556, -0.206)) and increased nitrogenous compounds in the water column (HDI: (0.097, 0.450)). The grazing submodel combined information from cow paddy density counts and disturbance classifications to generate values of latent grazing intensity for each site (Figure 3.4). As

we expected based on field observations, many sites experience moderate to high levels of grazing, while fewer experience very low levels of grazing.

Regionally, some species were far more variable in their distribution among parks than others. For example, although local factors tended to have minimal effects on the distribution of western toads (Figure 3.3), among-region variability was quite high for this species (Figure 3.7). In contrast, Pacific chorus frogs had low among-park variability, being nearly ubiquitous.

Most amphibian species were easier to detect during the first survey, probably due to metamorphosis occurring before or during the late-summer visit when evidence of successful breeding is no longer detectable (Figure 3.8). The inclusion of the visit number covariate (first vs. second) accounted for this discrepancy in the sense that species-specific covariate effects can still be recovered (Figure 3.6).

The shoreline vegetation submodel performed well, capturing the fact that some shorelines were either devoid or completely covered by vegetation, with variability between these two extremes (Figure 3.4). The latent N variable and observed log-transformed total dissolved N and NH_4^+ concentrations showed good fit, with increasing N concentrations in late summer compared to early summer (Figure 3.4). Graphical checks of independence assertions indicated no further causal pathways for inclusion (Figure 3.9).

3.9 Discussion

Our approach demonstrates the integration of SEM with occupancy modeling using a large-scale survey of pond breeding amphibians to gain a better understanding of the drivers of community composition. Importantly, this approach allows for explicit differentiation between observed data and underlying processes, accounting for a variety of measurement error models including the imperfect measurement process that gives rise to species detection data. We also embedded other types of observation models to account for more complicated likelihood functions, including the zero-one inflated beta distribution that was used to model observations of shoreline vegetation cover. With this approach, latent ecological processes hypothesized to drive occurrence can be

explicitly represented and their direct and indirect consequences formally evaluated.

The potential advantages of integrating occupancy and structural equation frameworks include: 1) developing more mechanistic approaches for understanding species distributions influenced by simultaneous related processes; 2) inheriting a formal method for evaluating potential outcomes that would result from management interventions; and 3) clarifying hypotheses by requiring that assumptions be represented in a formal causal model. Unlike purely associational methods, the causal assumptions embedded within structural equation models facilitate unique predictions that can be used to answer applied management questions.

Well-specified SEMs yield unique management insights based on “do” operators, which can estimate the consequences of targeted interventions from effect decompositions (Pearl, 1998). “Do” operators could be used to predict changes in occupancy following a management intervention that reduces grazing intensity or removes fish. Indeed, this was one of the original arguments for the development of structural equation models in the mid-20th century (Shipley, 2002). Causal inferences drawn from such analyses can be no better than the validity of the assumptions used to construct the model, however. Processes that drive the species distributions can be represented best when there is substantive knowledge to construct such models. This brings up two issues: how mechanistic is “mechanistic enough”, and how valid are the assumed causal relationships? First, it would be unreasonable to expect to construct an infinitely accurate and specific causal model for the occurrence of any species (Shipley, 2002). However, an adequate model may capture the most important drivers of occurrence and provide “good-enough” predictions for management interventions. This point underlines the role of interventional tests of such models - if the true outcome of management intervention deviates from anticipated effects, then the validity of the underlying causal model must be questioned.

Although not a substitute for interventions, tests of model fit to observational data can also be useful for identifying potentially missing pathways based on independence relationships among variables in the causal network. Here, we employed graphical checks of independence assumptions for observed variables, but better methods could be developed. Information theoretic and

d-separation tests exist for path analytic models, and other indices of model fit have long been used for traditional linear structural equation models that assume multivariate normality of errors (Grace, 2006; Shipley, 2013). However, methods for evaluating occupancy model fit are still relatively new and mostly rely on out-of-sample data (Zipkin et al., 2012) or bootstrapping (MacKenzie and Bailey, 2004). Developing methods to evaluate the fit of hierarchical occupancy-type structural equation models with non-normal indicators and binary, partially observed latent variables is a non-trivial task, but one that would increase the utility of this method. Advances in the study of probabilistic graphical models may provide solutions, but such advances are only just beginning to be applied by ecologists (Koller and Friedman, 2009; Grace et al., 2012). Aside from d-separation tests, management experiments provide another way to test the predictive power of these types of models.

3.10 Management implications

Our results have direct relevance to the management of threatened amphibian populations within lowland wetlands in California, particularly when land managers are faced with multiple potential challenges simultaneously. Livestock grazing, which is common throughout the western United States (Fleischner, 1994), has been a topic of uncertainty with regard to amphibian conservation. While several studies indicate negative overall effects of grazing on populations of specific amphibian species (Knutson et al., 2004; Schmutzler et al., 2008), others indicate the potential for positive effects of grazing on diversity and the perseverance of native communities (Marty, 2005).

Based on our analysis, livestock grazing in the Bay Area of California had minimal effects on the occurrence of six amphibian species, two of which are native species of conservation concern. This finding suggests that current grazing levels employed on these parks may be compatible with management aimed at conserving threatened amphibians. However, we did not explore the effects of grazing on amphibian abundance or temporal dynamics such as persistence and colonization. Such approaches may reveal effects of grazing that were not seen in this study.

Most ponds in this data set were initially constructed to serve as watering sites for livestock.

Because most original natural wetlands in California have been destroyed for agriculture and development, particularly in the Central Valley (Garone, 2011), such livestock ponds may now serve as vital habitat refuges for declining species. This trend is particularly important for the California tiger salamander and California red-legged frog, which are the focus of considerable conservation efforts due largely to habitat destruction (Lannoo, 2005).

In sharp contrast to the minimal effects of grazing, we found strong negative effects of nonnative fish on native amphibian occupancy for all species except western toads and nonnative bullfrogs. This finding supports a large body of research showing that native pond-breeding amphibians that lack evolutionary history with fishes are unlikely to persist once predatory fish have been introduced to a breeding site. The most common fish species at our field sites were mosquitofish and centrarchids (bass, bluegill and other sunfish), which are native to the eastern United States. These species likely prevent amphibian reproduction through direct predation on multiple life stages of native amphibians. The lack of an effect on toads is consistent with prior work showing that toxicity and schooling behavior of toad larvae provides resistance to predation by fish (Kruse and Stone, 1984; Welsh et al., 2006). Bullfrogs, which require permanent water bodies to complete metamorphosis, have co-evolved with fish in their native range, are unpalatable, and are not strongly affected by fish presence (Walters, 1975; Kruse and Francis, 1977; Szuroczki and Richardson, 2011).

Taken together, our results indicate that land management strategies should prioritize removal of nonnative fish rather than limitation of livestock grazing. Fish removal via pond draining has been shown to be effective in restoring populations of threatened amphibians within the study region (Alvarez et al., 2003). Considering the needs of interests groups involved and the prevalence of grazing, this strategy is perhaps more feasible than dramatically limiting access of livestock to ponds, which may have unintended side effects including elimination of pond breeding habitat due to overgrowth by vegetation. While more research is needed to evaluate both temporal effects of grazing, and how grazing affects amphibian abundance, our analysis indicates that fish introductions have more severe impacts than cattle grazing.

3.10.1 Conclusion

Structural equation modeling provides a framework to evaluate why species occur in some areas and not others, allowing a more direct confrontation of some of the most fundamental questions in ecology. Occupancy modeling provides a solution to the problem of imperfect detection and can account for many different processes giving rise to detection data. Combining these approaches provides a means to evaluate complex causal processes driving occurrence while accounting for false-absences in empirical occurrence data. This framework potentially facilitates deeper insights into biological processes, making a clear separation between imperfectly observed data and underlying states. From a pragmatic standpoint, it is clearly advantageous to be able to represent both direct and indirect determinants of species occurrence rather than being limited to treating covariates as being independent. This approach also provides the practical advantage of inheriting a suite of methods to anticipate the effects of management interventions (“do” operators), and to account for many different sampling schemes in the occupancy modeling literature. Future extensions of this method could deepen connections between SEM and other classes of occupancy models, including dynamic multi-year occupancy models, models of abundance such as N-mixture models, Dail-Madsen models, their multi-species extensions, and spatial models (Royle, 2004; Royle and Kéry, 2007; Dail and Madsen, 2011; Lamb et al., 2014; Dorazio and Connor, 2014).

The full value of this approach may be most apparent, and perhaps palatable to individuals unfamiliar with SEM, when coupled with controlled field experiments, where the validity of the underlying causal model may be tested directly. Compared to fields like sociology and economics, ecology as a field is perhaps in a unique position to reap benefits from SEM because it is both complex and relatively amenable to experimental manipulation.

Figure 3.1: Conceptual model showing influences on amphibian breeding success. Nodes represent concepts of interest, and directional arrows represent causal effects (e.g. grazing intensity affects shoreline vegetation)

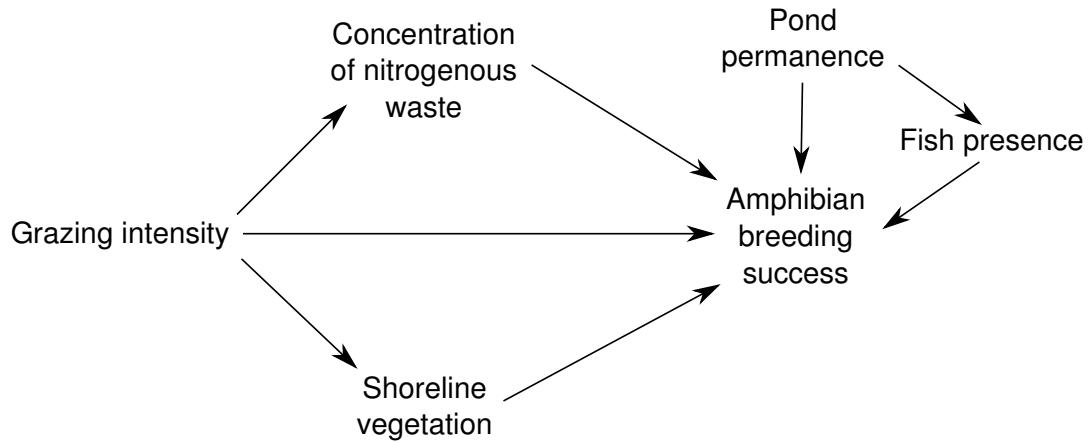


Figure 3.2: Directed acyclic graph illustrating relationships between unknown quantities (circles) and observed indicator variables (rectangles).

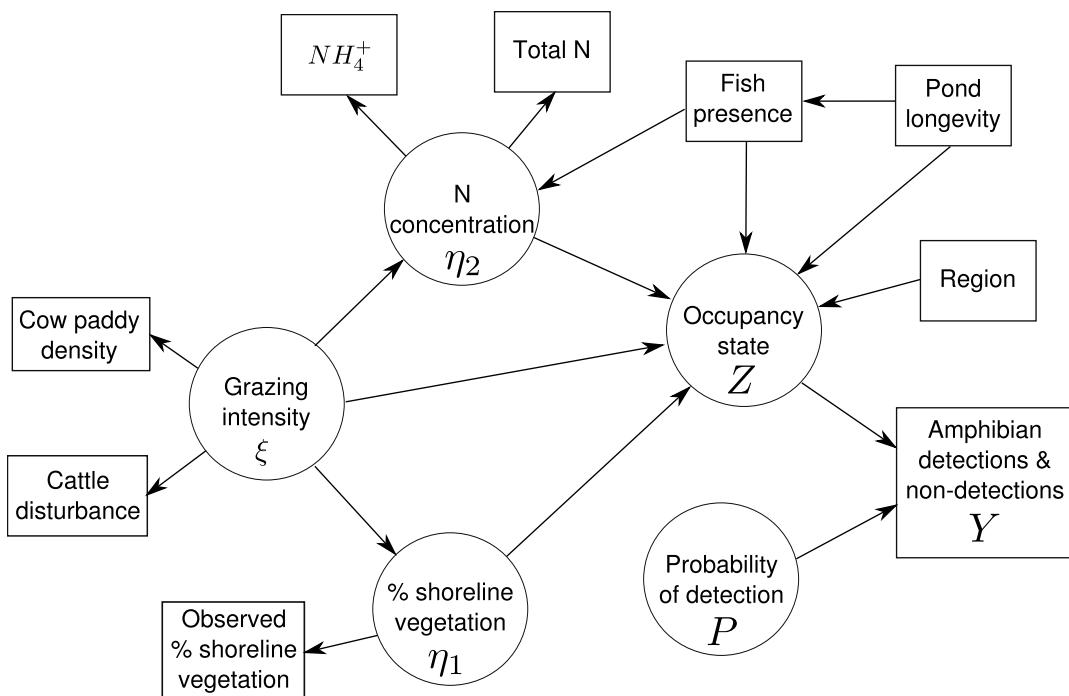


Figure 3.3: Estimated direct effects on occurrence probabilities for each species. Black corresponds to parameters for which HDIs excluded zero, and grey corresponds to HDIs including zero

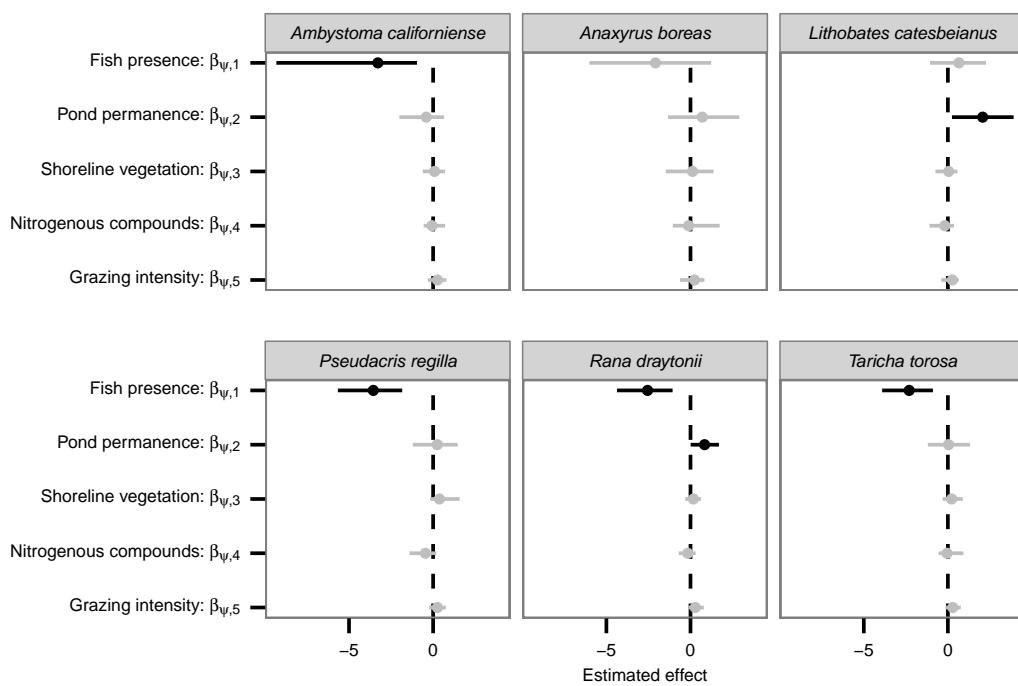


Figure 3.4: Fit of the submodels to the observed indicator variables. Shaded regions encompass the 95% HDI, with observed data shown as jittered points (x-axis values represent posterior medians).

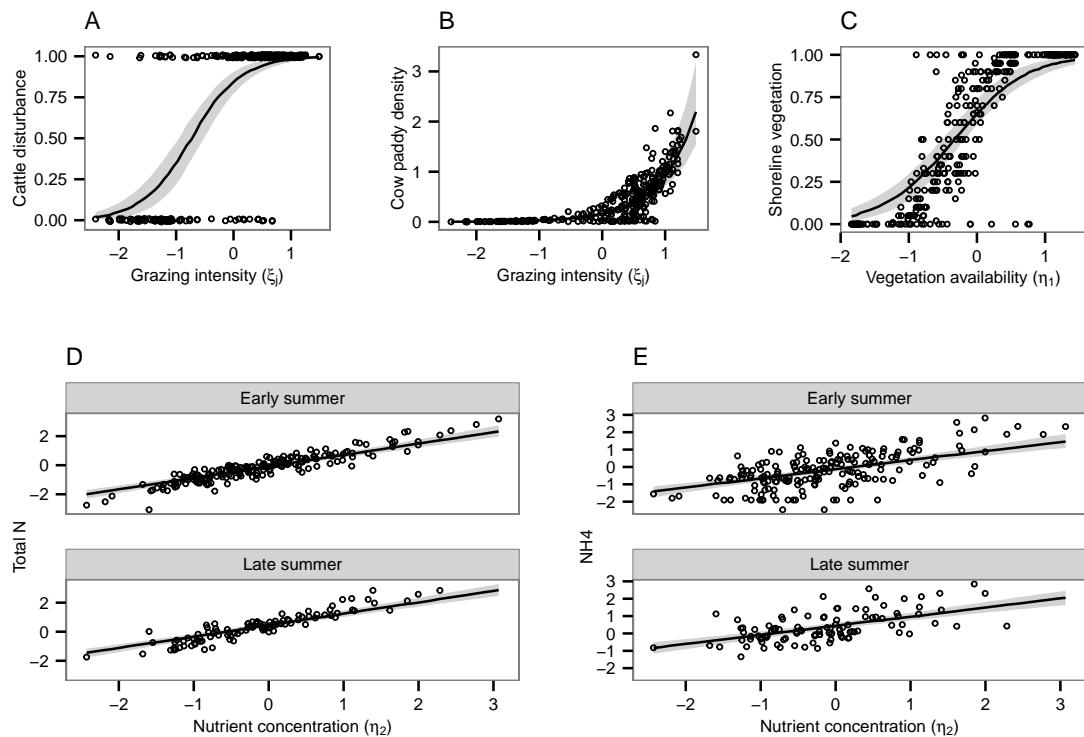


Figure 3.5: Results from a parameter recovery simulation in which >100 data sets were simulated from the model, and parameters were estimated. Percent coverage represents the percentage of highest posterior density intervals (HDIs) containing the true value for each parameter.

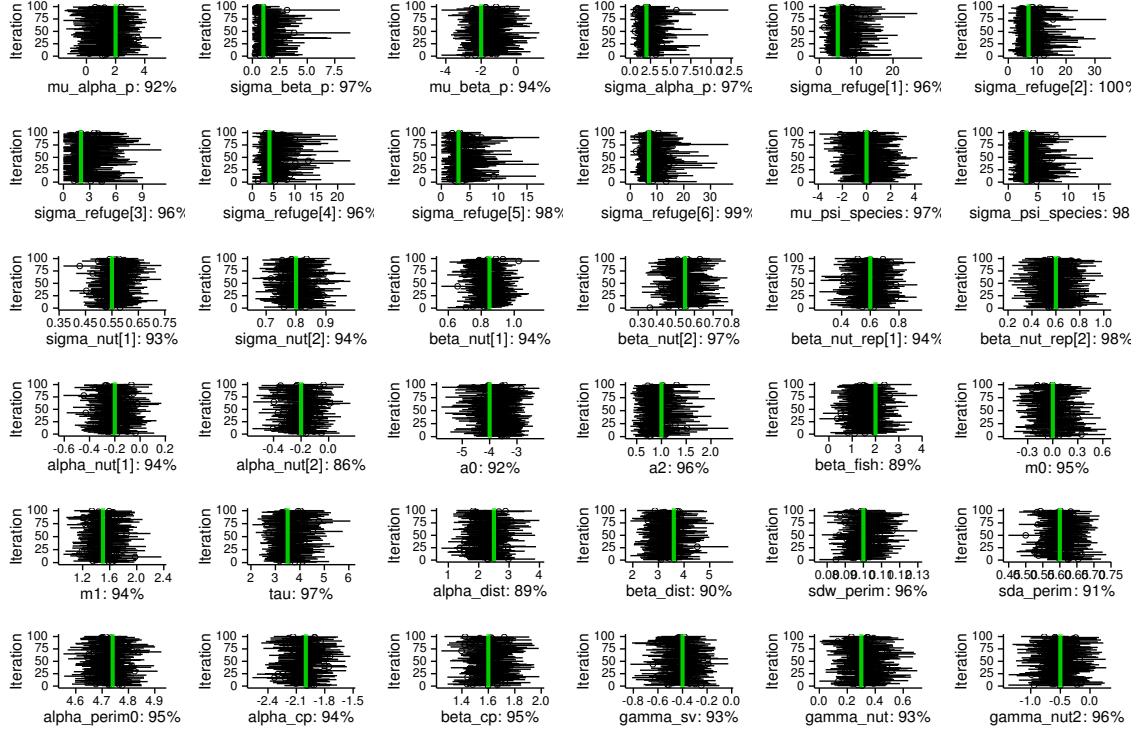


Figure 3.6: Recovery of β_ψ terms from the simulated 100 data sets. Red indicates parameter estimates for which the HDI was entirely negative, green indicates HDIs overlapping zero, and blue indicates HDIs that were entirely positive.

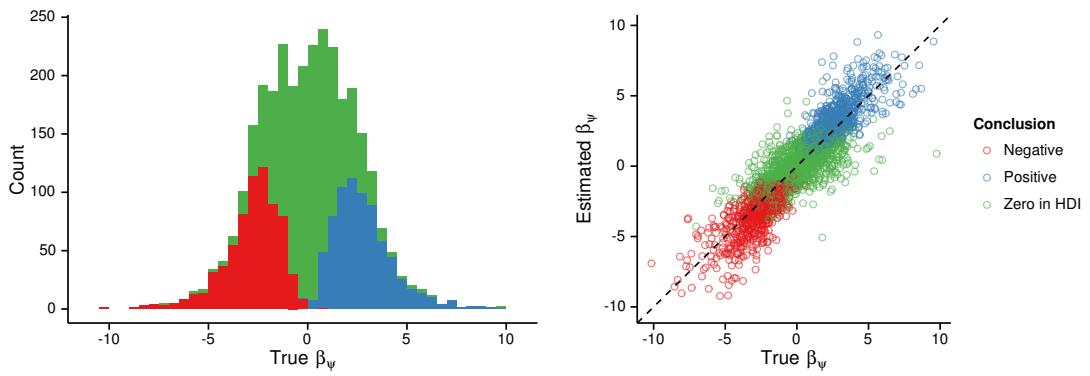


Figure 3.7: Estimated among-refuge standard deviations in occupancy probabilities for each species.

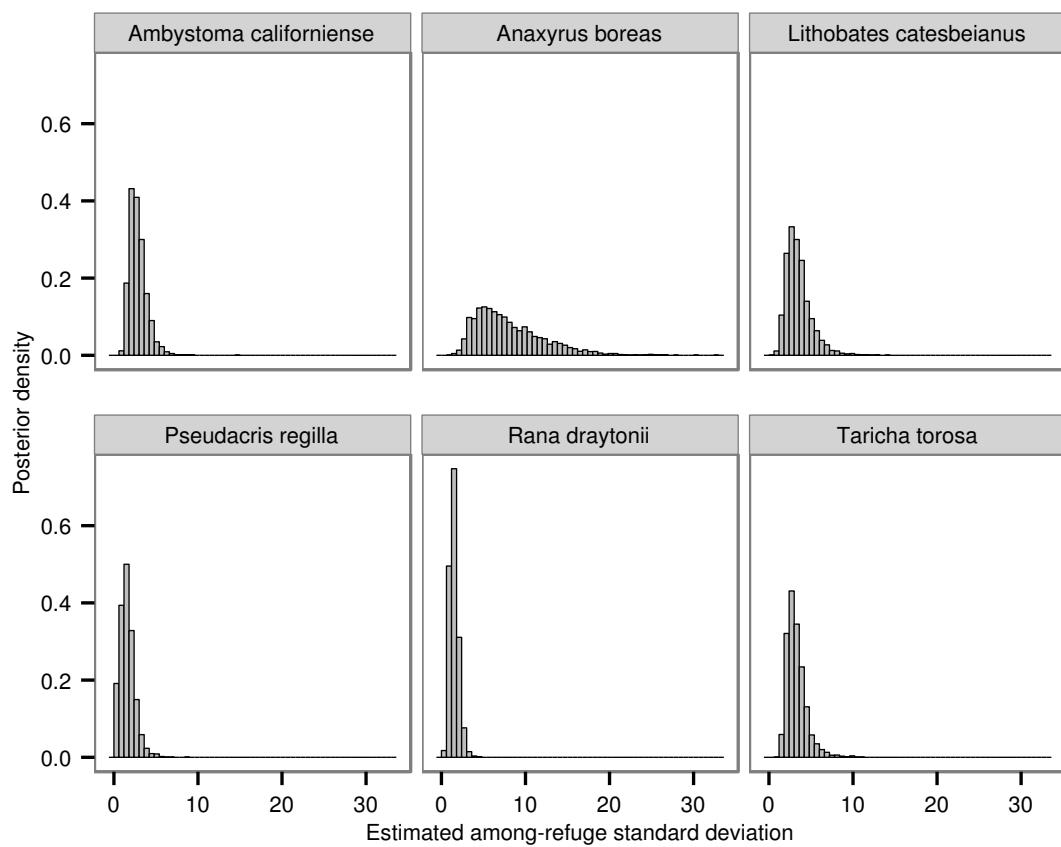


Figure 3.8: Posterior estimates of detection probabilities for each species in early and late summer. The interval represents the 95% HDI, and the points are placed at the posterior medians.

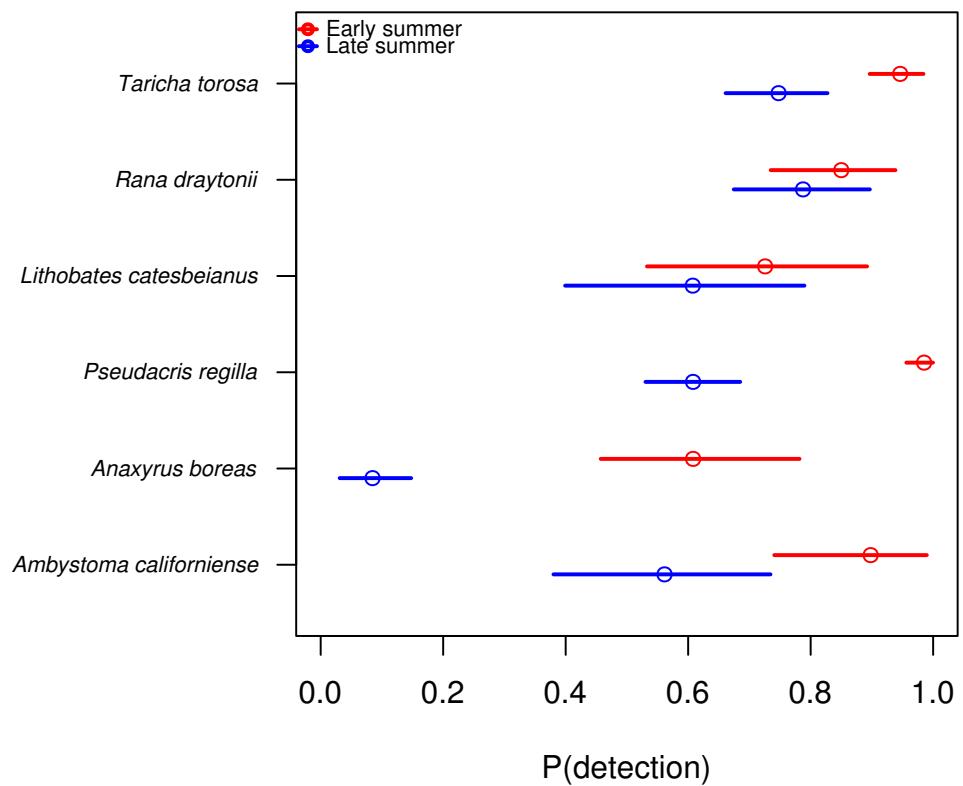
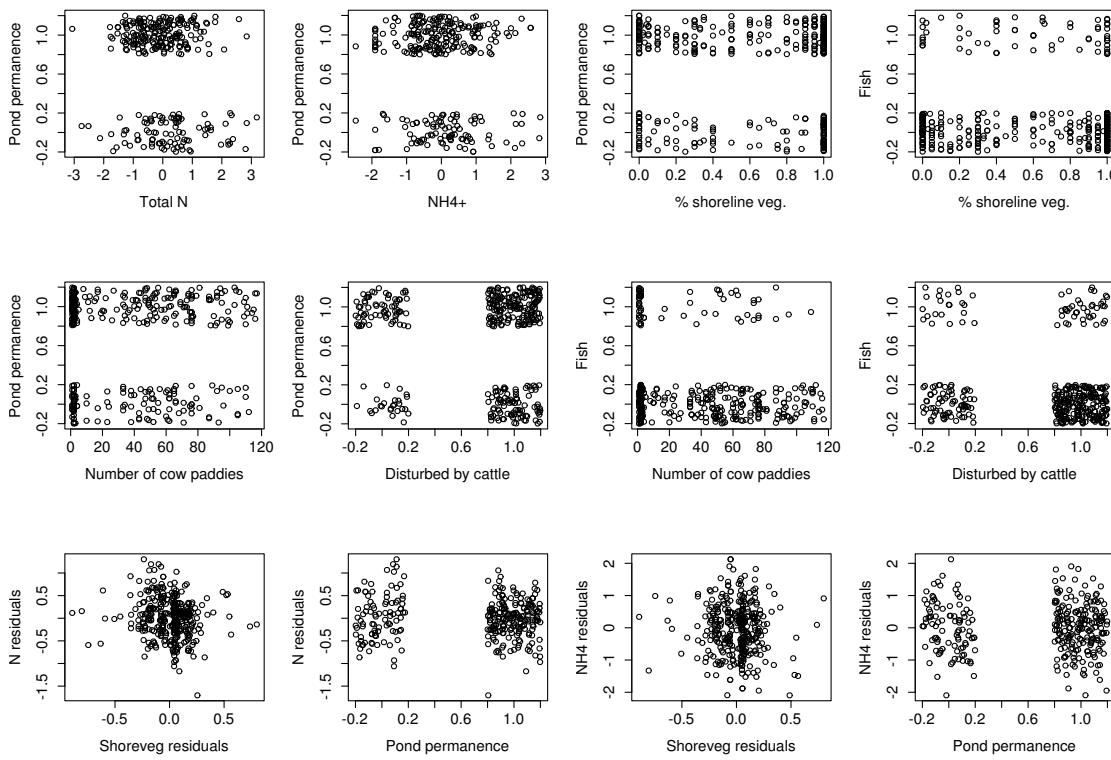


Figure 3.9: Graphical checks of independence assumptions implied by our model, used to check for missing causal pathways.



Chapter 4

Multilevel models for the distribution of hosts and symbionts

MAXWELL B. JOSEPH, WILLIAM STUTZ, PIETER T. J. JOHNSON

Symbiont occurrence is influenced by host occurrence and vice versa, which leads to correlations in host-symbiont distributions at multiple levels. Interactions between co-infecting symbionts within host individuals can cause correlations in the abundance of two symbiont species across individual hosts. Similarly, interactions between symbiont transmission and host population dynamics can drive correlations between symbiont and host abundance across habitat patches. If ignored, these interactions can confound estimated responses of hosts and symbionts to other factors. Here, we present a general hierarchical modeling framework for distributions of hosts and symbionts, estimating correlations in host-symbiont distributions at the among-site, within-site, among-species, and among-individual levels. We present an empirical example from a multi-host multi-parasite system involving amphibians and their micro- and macroparasites. Amphibian hosts and their parasites were correlated at multiple levels of organization. Macroparasites often co-infected individual hosts, but rarely co-infected with the amphibian chytrid fungus. Such correlations may result from interactions among parasites and hosts, joint responses to environmental factors, or sampling bias. Joint host-symbiont models account for environmental constraints and species interactions while partitioning variance and dependence in abundance at multiple levels. This framework can be adapted to a wide variety of study systems and sampling designs.

4.1 Introduction

Symbiotic organisms - those that live with, in, or on free living hosts - play important roles in disease dynamics, food production, and host health (Bashan, 1998; Jones et al., 2008). However, host-symbiont interactions complicate efforts to explain symbiont occurrence and abundance for several reasons. First, symbiont distributions depend on host distributions. In the extreme, obligate symbionts cannot exist without hosts (Moran and Baumann, 2000). Symbionts also influence host distributions through effects on fitness and population dynamics (Ebert et al., 2000; Lloyd-Smith et al., 2005). Further complexity arises in systems with multi-host symbionts, and host individuals infected with multiple co-infecting symbionts. Symbionts occupying the same host individual can interact, such that one symbiont may directly affect the distribution of another symbiont at the individual level (Telfer et al., 2010). Useful models of symbiont occurrence and abundance should accommodate these bidirectional influences and the hierarchical nature of host-symbiont interactions (Mihaljevic, 2012).

Multilevel modeling provides a promising avenue to understand patterns in host and symbiont abundance at different levels of biological organization (Gelman and Hill, 2007). A general host-symbiont modeling framework must be multivariate: any interaction between a host and a symbiont involves at least two species. Further, useful methods should make use of observable host and symbiont data which often consist of discrete counts, but may also include binary measurements of habitat use or continuous measures of density. Continuous and discrete multivariate observations can be modeled by combining univariate distributions with multivariate linear predictors, leading to a multivariate probit for binary data, multivariate Poisson for counts, and multivariate lognormal for continuous positive observations (Ashford and Sowden, 1970; Aitchison, 1982; Aitchison and Ho, 1989). Such models are increasingly being used to model distributions of free-living species while accounting for species interactions (Wisz et al., 2013; Clark et al., 2014; Pollock et al., 2014; Warton et al., 2015).

Here we expand upon existing methods to develop a hierarchical, multivariate framework for

modeling host and symbiont distributions that accounts for multiple levels of correlation, level-specific covariates, and flexible likelihood specifications. We begin by outlining the general features and logic of this approach. We then present an empirical case study of amphibian hosts and their parasites, revealing correlation among species at multiple levels and demonstrating the types of insights gained in practice. We conclude by discussing limitations and potential extensions.

4.2 Methods

While ecologists often seek to estimate the effects of one species on another species, this requires strong causal assumptions when working with observational data (Pearl, 2000b). Instead, correlations in species abundance and occurrence - potentially resulting from species interactions - can be modeled as a proxy, helping to generate hypotheses about interactions that ideally can be pursued experimentally (Ovaskainen et al., 2010). Due to the hierarchical nature of host-symbiont interactions, these correlations can occur at multiple levels (Mideo et al., 2008). Symbionts may be correlated at the level of host individuals, positively if two symbiont species often co-infect hosts (e.g., Puoti et al., 2002). Symbionts may also be correlated at the level of host species, positively if two symbionts tend to infect the same species (e.g., Johnson, 2012). Hosts and symbionts might also be correlated within and among spatial locations (hereafter "sites"). While such correlations can arise through species interactions, they can also emerge from simultaneous responses to extrinsic factors or sampling bias. These alternative drivers of correlations are not guaranteed to be differentiable from observational data alone (Pearl, 2000b; Dorazio and Connor, 2014), emphasizing the importance of methods that limit causal assumptions.

We consider a landscape with discrete habitat patches (sites) containing multiple species of hosts and symbionts. At each site, replicate surveys are conducted to measure host density, and symbiont abundance is observed by sampling individual hosts. We assume each host species $h = 1, \dots, H$ is present or absent at each site $i = 1, \dots, N$, with occurrence constant across surveys. If they are present, they have some non-zero site-level average density μ_{ih} . The probability of occurrence ψ_{ih} and expected density within a site are assumed to be proportional (He and Gaston,

2003). Hosts may be present at a site but unobserved (MacKenzie et al., 2002). Conditional on occurrence, the probability of detection increases with density (Royle et al., 2005). In other words, sites favoring high density are commonly occupied, and abundant hosts are easier to detect than rare hosts. At site i , $J_i > 1$ repeat surveys are conducted, leading to the following likelihood or sampling distribution for host abundance observations:

$$y_{ih} \sim \begin{cases} \psi_{ih} \prod_{j=1}^{J_i} f(y_{ihj} | \theta_{ihj}), & \sum_{j=1}^{J_i} y_{ihj} > 0 \\ \psi_{ih} \prod_{j=1}^{J_i} f(0 | \theta_{ihj}) + 1 - \psi_{ih}, & \text{otherwise} \end{cases} \quad (\text{eqn 1})$$

Where y_{ih} is a vector of length J_i with elements consisting of abundance measurements (e.g., counts) of species h at site i in each survey. This is a mixture model with components representing cases in which species h is present or absent from site i with probabilities ψ_{ih} and $1 - \psi_{ih}$, respectively. Further, $f(y | \theta_{ihj})$ is a probability density or mass function with parameter(s) θ_{ihj} potentially varying across sites, species, and surveys (Ver Hoef and Boveng, 2007). If species h is not observed at site i , then it was absent with probability $1 - \psi_{ih}$ or present but unobserved with probability $\psi_{ih} \prod_{j=1}^{J_i} f(0 | \theta_{ihj})$. False absences are more likely for species with low densities and those highly aggregated within sites. For simplicity we assume that detection implies species presence, but a likelihood could be specified to account for false positives (Royle and Link, 2006).

We assume that the occupancy probability of species h at site i increases with expected density μ_{ih} as follows (He et al., 2002):

$$\text{logit}(\psi_{ih}) = \gamma_{0h} + \gamma_{1h} \log(\mu_{ih}) \quad (\text{eqn 2})$$

Here, γ_{0h} is the probability of host species h occurring at site i on a logit scale when the mean density is one individual per unit area of habitat (e.g. per square meter), and γ_{1h} is a parameter that describes the scaling between expected density and the probability of occupancy, which we expect to be positive. This occupancy submodel could also include covariates such as habitat area.

Symbiont species $s = 1, \dots, S$ are present or absent at each site. At site i , K_i host individuals

are sampled and their infections quantified. Non-detection of symbiont s at site i can result from true absence or failure to sample an infected host, and sites favoring high symbiont abundance are more likely to be occupied:

$$y_{is} \sim \begin{cases} \psi_{is} \prod_{k=1}^{K_i} f(y_{isk} | \theta_{isk}), & \sum_{k=1}^{K_i} y_{isk} > 0 \\ \psi_{is} \prod_{k=1}^{K_i} f(0 | \theta_{isk}) + 1 - \psi_{is}, & \text{otherwise} \end{cases} \quad (\text{eqn 3})$$

$$\text{logit}(\psi_{is}) = \gamma_{0s} + \gamma_{1s} \log(\mu_{is}) \quad (\text{eqn 4})$$

Every host and symbiont species has a site-level mean density, and these densities may be correlated e.g., if an abundant reservoir host increases infection in other hosts (Ashford, 2003). Species have some among-site variance in their abundances, and these variance parameters may differ across species. Species that are always at low or high abundance will have low variance, and species that are abundant in some sites, and absent from others will have higher variance. These correlation and variance parameters are used to construct a covariance matrix Σ_{site} with elements $\rho_{mn}\sigma_m\sigma_n$ in the m^{th} row, n^{th} column, where ρ_{mn} is the correlation between species m and n , and σ_m is the among site standard deviation for species m . Each site has a random effect vector α_i of length $H + S$: $\alpha_i \sim N_{H+S}(\mathbf{0}, \Sigma_{site})$, where $N_d(\mathbf{0}, \Sigma)$ represents a multivariate normal distribution with dimension d , mean vector $\mathbf{0}$, and covariance matrix Σ .

Within sites, hosts and symbiont density can vary among survey locations. Uniformly distributed species have low variance, and spatially aggregated species have high variance. Species are correlated within sites if they tend to be observed together in the same surveys more or less often than expected by chance, for example. We can represent these survey level correlations and variance parameters in a covariance matrix Σ_{survey} , which gives rise to $J_{tot} = \sum_i j_i$ survey level random effect vectors α_j , each with length $H + S$: $\alpha_j \sim N_{H+S}(\mathbf{0}, \Sigma_{survey})$. Random effects may be adapted to alternative sampling designs. For instance, if hosts are sampled for symbionts independently from host density surveys, then symbionts are not associated with particular surveys and the survey-level random effects may instead have dimension H .

Differences in overall mean abundance are represented with a host species specific random effect α_{0h} which is univariate normally distributed around a community mean, with among species variance. Together, these random effects contribute to the expected number of individual hosts of species h detected in a survey j at site i , here with a log-link:

$$\log(\mu_{ihj}) = \alpha_{0h} + \alpha_{jh} + \alpha_{ih} \quad (\text{eqn 5})$$

Depending on survey design, this expectation might include an offset that accounts for among-survey variation in sampling time intervals or area (Gelman and Hill, 2007).

The expected density of symbionts also includes an intercept α_{0s} and elements from the site-level and survey-level random effects. However, because of the nature of host-symbiont interactions, symbionts have the potential for correlation at additional levels. Specifically, symbionts may be correlated at the individual host level, e.g., if two symbionts commonly co-infect host individuals. We represent these host individual differences with $K_{tot} = \sum_i K_i$ multivariate normal random effects with mean zero and covariance matrix Σ_{indiv} including correlation terms and symbiont species specific variance terms representing how variable host individuals are in their infection abundances: $\boldsymbol{\alpha}_k \sim N_S(\mathbf{0}, \Sigma_{indiv})$.

Finally, hosts may vary in their symbiont infection abundances at the species level. This variation may be correlated if two host species are functionally alike, e.g., they tend to be similarly susceptible to infection across a range of symbiont species. To allow for species level variation we consider $h = 1, \dots, H$ multivariate normal random vectors, each with S elements: $\boldsymbol{\alpha}_h \sim N_S(\mathbf{0}, \Sigma_{species})$.

Together, these random effects contribute to the expected infection load of symbiont s at site i in host individual k of species h sampled in survey j :

$$\log(\mu_{isk}) = \alpha_{0s} + \alpha_{is} + \alpha_{j[k]s} + \alpha_{hks} + \alpha_{ks} \quad (\text{eqn 6})$$

If host sampling for symbionts occurs separately from host abundance surveys, then sampled hosts are not associated with surveys, simplifying the random effects:

$$\log(\mu_{isk}) = \alpha_{0s} + \alpha_{is} + \alpha_{h_k s} + \alpha_{ks} \quad (\text{eqn 7})$$

4.3 Case study: amphibian communities and their parasites

Amphibians in the San Francisco Bay Area of California are infected with a diverse suite of parasites, including macroparasitic helminth worms (*Ribeiroia ondatrae* Looss, 1907, *Echinostoma* sp., *Cephalogonimus* sp., *Alaria* sp.), and microparasites such as *Ranavirus* sp. and the amphibian chytrid fungus *Batrachochytrium dendrobatidis*, Longcore, Pessier & D.K. Nichols (1999), hereafter referred to as *Bd*.

Five amphibian hosts comprise the majority of non-threatened (available for sampling) amphibian species: the Pacific chorus frog *Pseudacris regilla* (Baird & Girard, 1852), California newt *Taricha torosa* (Rathke, in Eschscholtz, 1833), rough-skinned newt *Taricha granulosa* (Skilton, 1849), western toad *Anaxyrus boreas* Baird & Girard, 1852, and the non-native American bullfrog *Lithobates catesbeianus* (Shaw, 1802) (Johnson et al., 2013). Previous studies in this system have revealed correlations between parasites at the host individual and site levels (Johnson and Buller, 2011; Hoverman et al., 2013).

In 2013, field crews visited 87 wetland sites in Contra Costa, Alameda, and Santa Clara counties. At each site, crews conducted dip net sweep surveys ($\bar{J}_i = 9.9$, standard deviation (s_{J_i}) = 1.2, range = [2, 15], $J_{tot} = 914$) to quantify amphibian density, recording the numbers and species identities of all amphibians observed. Crews collected hosts at each site to quantify parasite infections ($\bar{K}_i = 17.8$, $s_{K_i} = 12.7$, range = [1, 82], $K_{tot} = 1550$), and these collection events were separate from the sweep surveys. Collected hosts were larval or recently metamorphosed. We assessed macroparasite infection abundance via dissection (Johnson et al., 2013), and infection loads of Bd and *Ranavirus* using quantitative polymerase chain reaction of skin swabs and organ tissue, respectively (Hyatt et al., 2007; Hoverman et al., 2010).

4.4 Parameter estimation

We used a Bayesian approach to estimate parameters, combining prior information with a Poisson likelihood to generate a posterior distribution for unknown quantities. We simulated samples from the posterior using Markov chain Monte Carlo (MCMC) sampling in the probabilistic programming language Stan (Hoffman and Gelman, 2014). All data and code required to reproduce the analysis is available in the code supplement.

4.5 Results

We uncovered correlations between hosts and parasites at every level in the model. At the site level, we detected multiple correlations between hosts and parasites (Figure 4.1). Sites with high densities of Pacific chorus frogs had high densities of California newts and western toads, possibly due to similar habitat requirements (Joseph et al., 2015). Sites with high densities of chorus frogs had higher Bd infection loads, consistent with this species' role as a reservoir host (Reeder et al., 2012). Sites with high levels of infection of *Cephalogonimus* tended to have lower levels of infection with Bd. Macroparasites were positively correlated across sites, probably due to availability of planorbid snails that release macroparasite infective stages (cercariae), and deposition of parasite eggs in feces of carnivorous definitive hosts.

Within sites at the survey level, California newts correlated positively with rough-skinned newts and Pacific chorus frogs (Figure 4.2). These correlations imply that these species tend to be co-aggregated within sites, potentially due to similar microhabitat preferences.

At the host species level, among-parasite correlations were estimated with low precision as we would expect when trying to estimate a correlation with five points (host species). However, some posteriors leaned toward positive correlations e.g., between Bd and *Alaria* (Figure 4.3). This was driven by high infection abundances of most parasites in Pacific chorus frogs, consistent with these fast-lived hosts investing little in parasite defense (Johnson, 2012). More host species are needed to make reliable inference at this level.

At the individual host level macroparasite loads correlated positively, so that if an individual was heavily infected with one macroparasite, it was more likely to be heavily infected with other macroparasites (Figure 4.4). These positive correlations can occur despite negative within-host interactions (Johnson and Hoverman, 2012). For instance *Ribeiroia* and *Echinostoma* both have negative effects on the persistence of one another within host individuals, and the positive correlation may result from these parasites having similar niche requirements and host preferences (Johnson and Buller, 2011). In contrast, Bd correlated negatively with three macroparasites: *Ribeiroia*, *Alaria*, and *Echinostoma*. Parasite interactions could drive these correlations or they could result from confounding variables. For example, host age increases cumulative exposure, confounding inference on parasite interactions derived from correlations. Such correlations may disappear after including the confounding trait as a covariate, contingent on the validity of the model with respect to the true latent processes (Pearl, 2000b). Last, correlations could arise from sampling bias (Berkson, 1946). For instance, if Bd or *Echinostoma* infection increases catchability, then these two parasites will correlate negatively in our sample even if they are not correlated within the population.

We partitioned variation in host and parasite abundance among model levels to better understand the relative strength of processes operating at different scales. This analysis aims to summarize the correlations and extra-Poisson variance induced by the random effects. We considered effective variance $V_e(X) := |\Sigma_X|^{1/d}$, the d -th root of the determinant of a covariance matrix Σ_X with dimension d , which represents the average scatter in any direction (Peña and Rodríguez, 2003). We also considered effective dependence $D_e(X) := 1 - |\mathbf{R}_X|^{1/d}$, where \mathbf{R}_X is a correlation matrix, which captures the stochastic dependence among species (Peña and Rodríguez, 2003). If species tend to be highly correlated, this parameter will be close to one. With no correlation among species, effective dependence is zero. Within-site, among-survey variation accounted for less variation in host abundance than among site random effects (Figure 4.5). For parasites, variation among host individuals exceeded among-site variation. This is striking, but consistent with the notion that parasites are overdispersed and aggregated among host individuals (Anderson and

May, 1978). Despite high variance, parasite abundance showed relatively low dependence at the individual level. Effective dependence was comparable across other model levels, which might be expected if species interactions and/or joint responses to covariates similarly influence patterns of co-aggregation at these levels.

4.6 Discussion

We presented a general hierarchical modeling framework to understand correlations and drivers of host and symbiont abundance. This builds upon existing multi-species abundance models and specifically extends a two symbiont abundance model by Stutz et al. *in review*, allowing for more than two species of symbionts, inclusion of hosts (any number of species), partially observed occurrence states, and greater flexibility in likelihood specification. Many host-symbiont distributions could be investigated with this method beyond host-parasite associations, including commensal and mutualistic symbionts of plants and animals.

This approach has been described primarily from a causally agnostic perspective, in which we are estimating unstructured correlations among species, but alternative approaches could be taken. If there is a known causal direction, e.g., in an experimental setting, one could extend this method to model the effect of host density on symbiont abundance rather than their correlation. Estimation of many covariance matrices is a rather data-hungry operation, particularly when the correlation parameters are free to vary independently. If less information were available, it may be advantageous to either include structure for the correlation parameters (e.g., Dorazio and Connor (2014)), or adopt a latent factor approach as recently described by Warton et al. (2015). In the context of host-symbiont models, latent factors could be used at multiple levels to account for unobserved site, host, and species level characteristics.

Another advantage of this joint modeling approach is the ability to decompose variation and dependence across multiple levels of organization. Effective variance and dependence may reflect the relative importance of processes at different levels of organization. For instance, we found variation among host species in parasite abundance comparable to variation among spatial

locations, both of which exceeded variation within sites. Generally, the contribution of model levels to effective variance will differ among study systems, and the ability to compare across levels should be valuable in determining how to begin model expansion. In our case study for instance, a logical next step would be inclusion of site and host individual level covariates.

Alternative likelihood functions, including those accounting for measurement error, can be readily combined with this method. Here we made use of a Poisson likelihood, but some situations may call for the use of zero-inflated probability distributions with support for all real positive values, such as a zero-inflated lognormal or gamma (Miller et al., 2012). This would allow for direct modeling of observations generated via quantitative polymerase chain reaction, typical of applications to viruses and bacteria, and environmental DNA of free-living species. Continuous distributions would circumvent the need to round values for use with Poisson or negative binomial distributions with integer support. Last, we have assumed that infections are detected without error, but a rich set of methods could be applied to account for error in this measurement process (Lachish et al., 2012; Miller et al., 2012).

We assumed that sites favoring high density are more likely to be occupied. However, if different processes drive species occurrence and abundance, then alternative occurrence submodels could be developed. In particular, spatial and temporal dependence may be useful for representing limits to species occurrence (Holt and Keitt, 2000). Future developments of this approach might prioritize inclusion of spatiotemporally explicit colonization dynamics that account for occupancy status of neighboring sites, habitat quality, and dispersal functions (Broms et al., 2015). These approaches will prove useful to understand how much of the spread of an invasive symbiont may be due to changes in the host distribution vs. changes in the symbiont distribution alone, with potential applications to the management of emerging infectious diseases (Mitchell et al., 2006).

Symbionts have received an increased appreciation over past decades as the field of disease ecology has gained momentum and as modern genetic methods have increased our ability to sample unculturable communities (Schrag and Wiener, 1995; Riesenfeld et al., 2004). However, the development of methods to understand the distribution of symbionts has not kept pace with devel-

opments in free living species (Bailey et al., 2014). The approach presented here draws upon these developments with the goal of producing a general approach that can be readily adapted to other host-symbiont systems. Simultaneously modeling hosts and their symbionts in this hierarchical framework provides a powerful method to dissect patterns of occurrence and abundance for free living and symbiotic organisms.

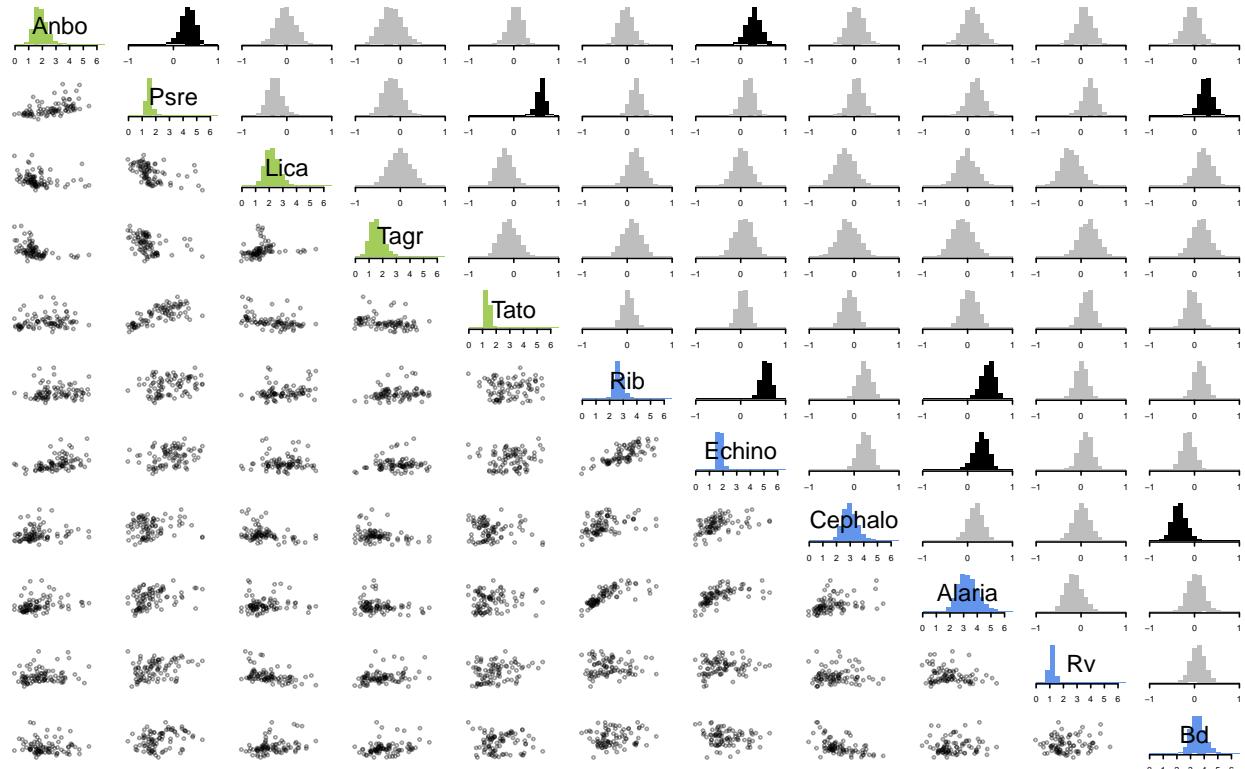


Figure 4.1: Site level variance covariance matrix and random effect posteriors. Diagonal elements display the among-site standard deviation in abundance for all host and parasite species (Anbo = *Anaxyrus boreas*, Psre = *Pseudacris regilla*, Lica = *Lithobates catesbeianus*, Tagr = *Taricha granulosa*, Tato = *Taricha torosa*, Rib = *Ribeiroia ondatrae*, Echino = *Echinostoma* sp., Cephalo = *Cephalogonimus* sp., Alaria = *Alaria* sp., Rv = *Ranavirus* sp., Bd = *Batrachochytrium dendrobatidis*). Green indicates hosts and blue parasites. Upper triangular elements show among-species correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate scatter plots of the posterior means of the site-level random effects corresponding to the intersection of the species in the rows and columns.

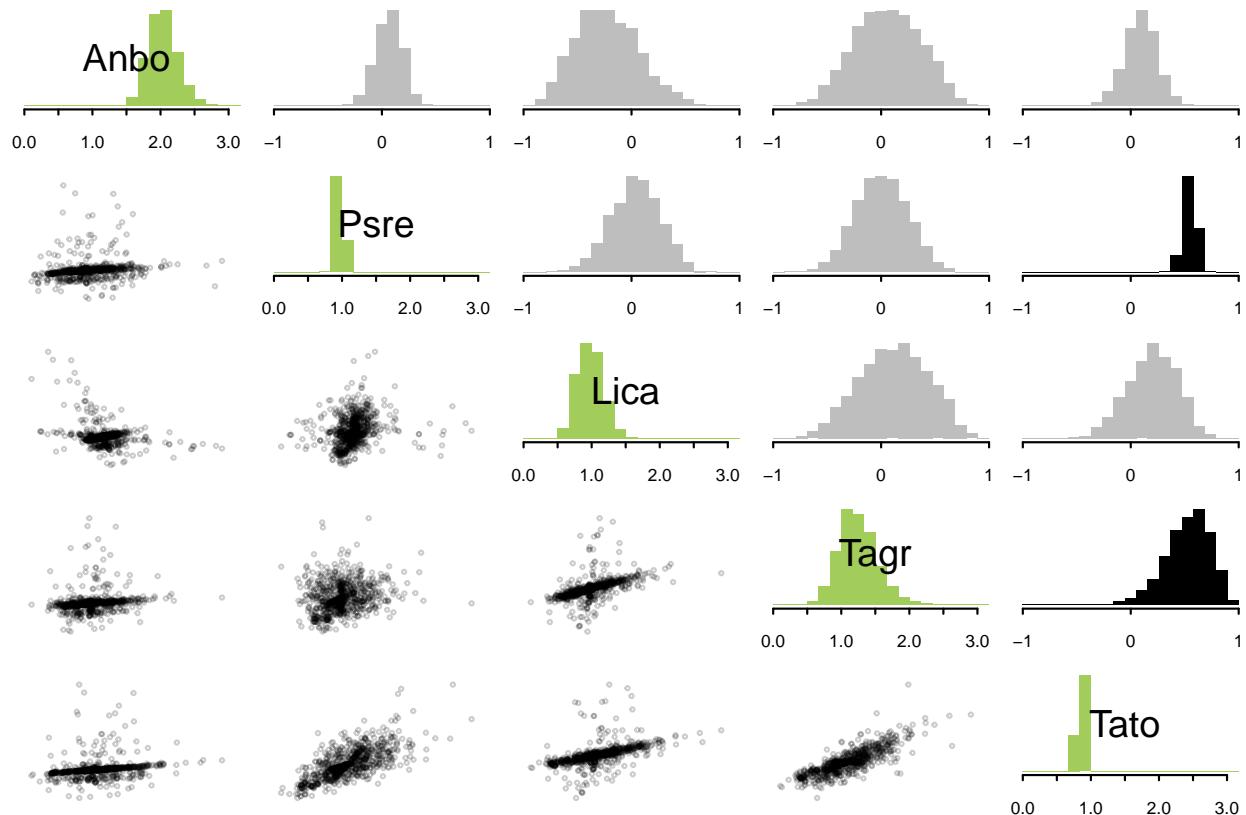


Figure 4.2: Survey level variance covariance matrix and random effect posteriors. Diagonal elements display the among-survey standard deviation in abundance for host species (Anbo = *Anaxyrus boreas*, Psre = *Pseudacris regilla*, Lica = *Lithobates catesbeianus*, Tagr = *Taricha granulosa*, Tato = *Taricha torosa*). Upper triangular elements show among-species correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate scatter plots of the posterior means of the survey-level random effects corresponding to the intersection of the species in the rows and columns.

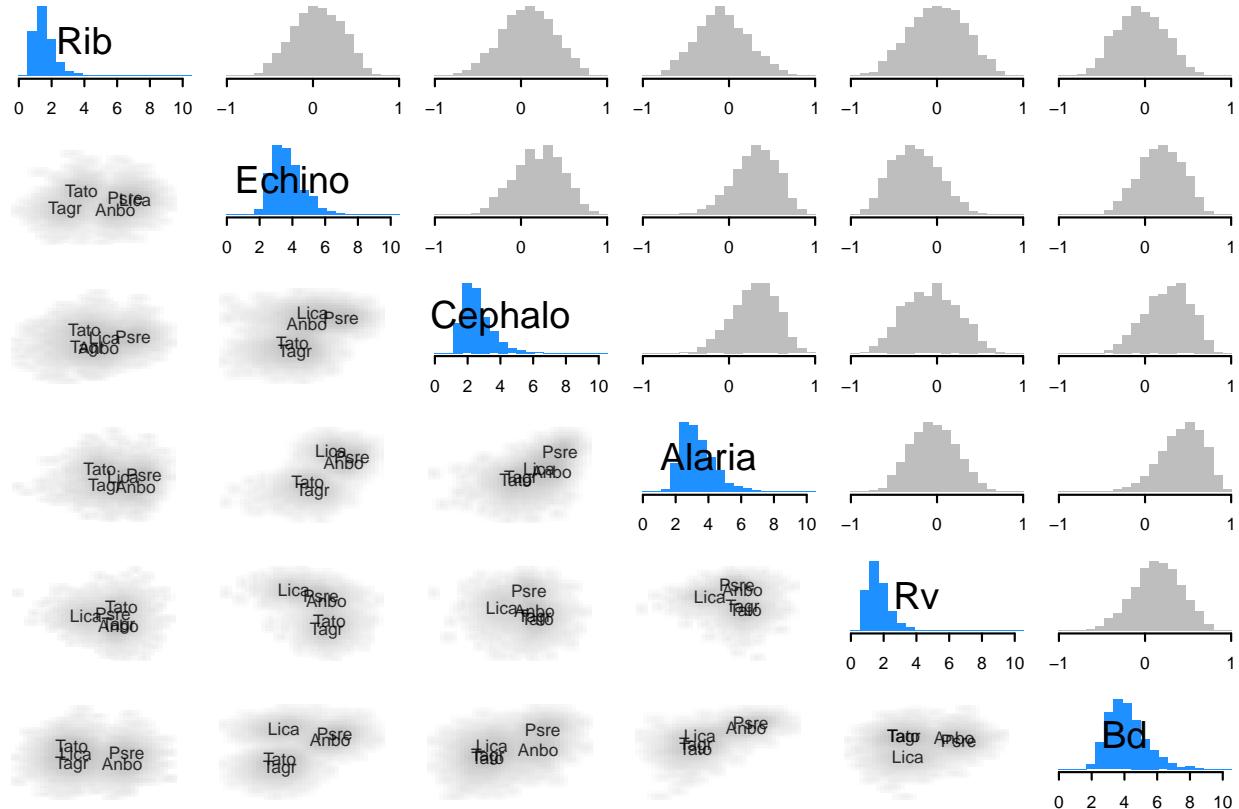


Figure 4.3: Host species level variance covariance matrix and random effect posteriors. Diagonal elements display the among host species standard deviation in abundance for parasite species (Rib = *Ribeiroia ondatrae*, Echino = *Echinostoma* sp., Cephalo = *Cephalogonimus* sp., Alaria = *Alaria* sp., Rv = *Ranavirus* sp., Bd = *Batrachochytrium dendrobatidis*). Upper triangular elements show among-species correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate smoothed scatter plots of species-level random effects, with host species codes printed at the posterior means (Anbo = *Anaxyrus boreas*, Psre = *Pseudacris regilla*, Lica = *Lithobates catesbeianus*, Tagr = *Taricha granulosa*, Tato = *Taricha torosa*).

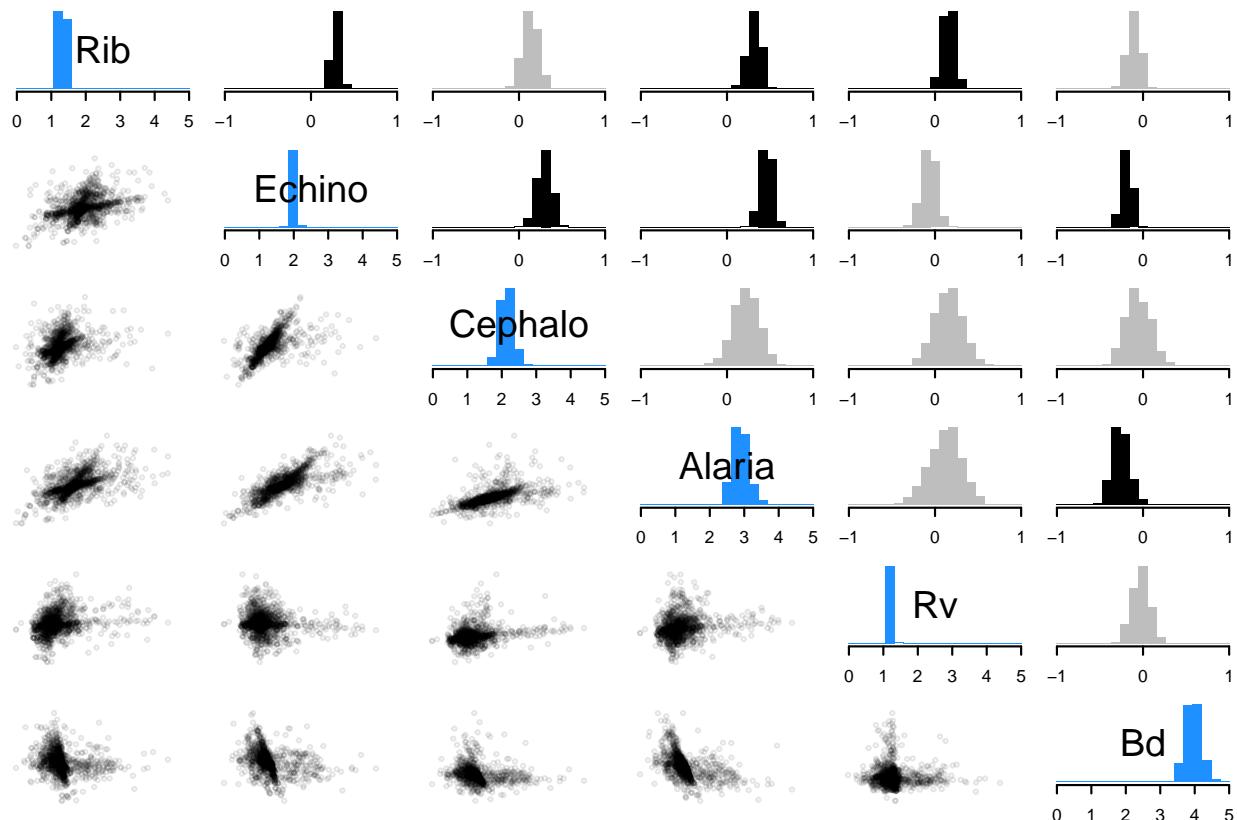


Figure 4.4: Individual level variance covariance matrix and random effect posteriors. Diagonal elements display the among host individual standard deviation in abundance for parasite species (Rib = *Ribeiroia ondatrae*, Echino = *Echinostoma* sp., Cephalo = *Cephalogonimus* sp., Alaria = *Alaria* sp., Rv = *Ranavirus* sp., Bd = *Batrachochytrium dendrobatidis*). Upper triangular elements show among-individual correlation parameters. Black indicates correlations that are probably positive or probably negative (95% of posterior probability mass greater than or less than zero); grey indicates otherwise. Lower triangular elements show bivariate scatter plots of the posterior means of the individual-level random effects corresponding to the intersection of the species in the rows and columns.

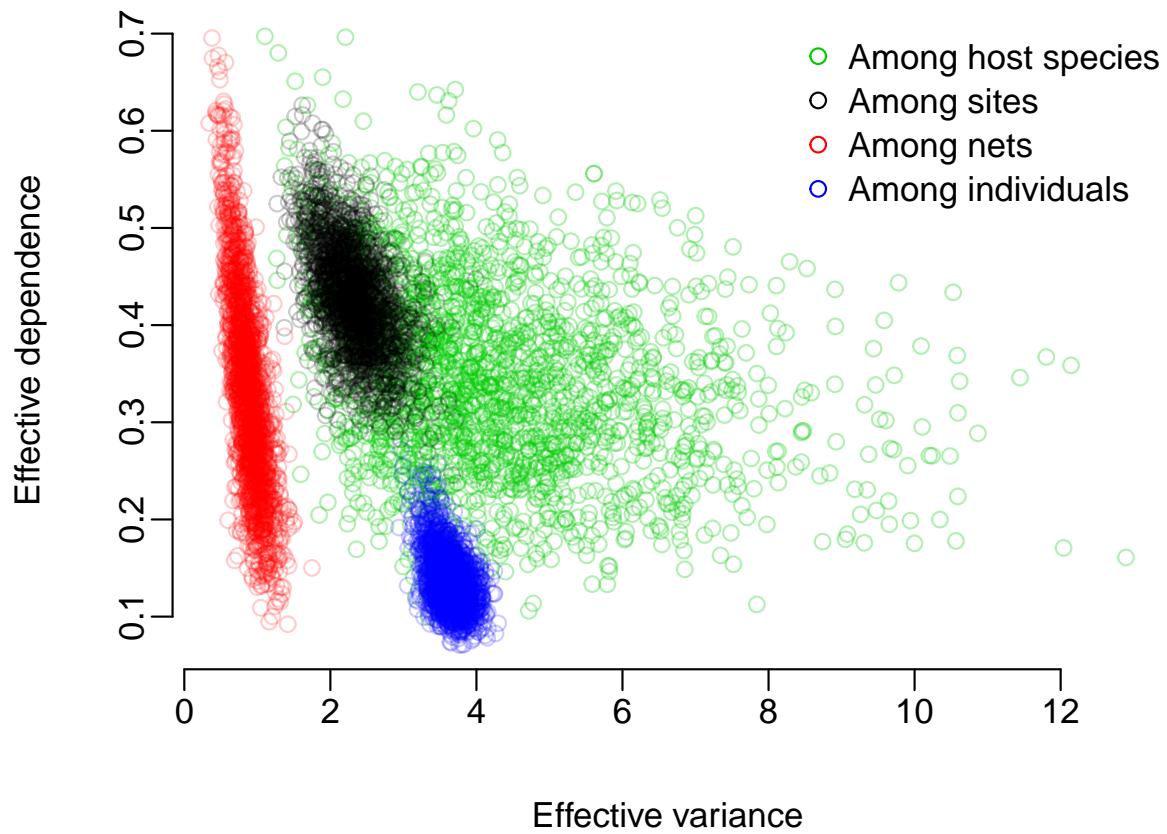


Figure 4.5: Bivariate posterior distributions of the effective variance and dependence for the multivariate random effects. Each point represents a simulated draw from the posterior. Effective variance measures the magnitude of spread in any direction of the random effects, and effective dependence measures the magnitude of among-species correlation.

Chapter 5

Species trait prediction via phylogenetic latent factors

Maxwell B. Joseph, Kevin B. Knight, Pieter T. J. Johnson

As trait-based approaches in ecology have gained momentum, efforts to assemble trait databases have increased the availability of trait data for many taxa. However, observations are often missing from these databases, and species with incomplete trait records may be excluded from analyses, thereby limiting the utility of trait-based approaches and potentially biasing subsequent inference. Here we develop a hierarchical Bayesian method to interpolate missing traits, exploiting correlations among traits and phylogenetic autocorrelation. The model maps observable traits to phylogenetically structured latent factors. We demonstrate this approach on a database of primate species traits with K-fold cross validation to approximate out-of-sample predictive power for models with varying latent factor dimensionality and phylogenetic correlation structure. The expected and predicted values were highly correlated with the held-out values ($r = 0.86$ and 0.79 , respectively), and 95% of the 95% prediction intervals contained the true trait value. This latent factor approach provides a means to interpolate missing observations in species level trait databases, particularly for poorly described species with well described evolutionary neighbors. We conclude with some promising applications for estimating extinction risk, along with methodological extensions to this latent factor approach.

5.1 Introduction

In recent decades, many species trait databases have been assembled to meet the needs of trait-based approaches, which rely on species traits to achieve a more quantitative understanding of ecological phenomena, rather than simply treating species as categorically different (Ackerly and Cornwell, 2007; Gross et al., 2009; Webb et al., 2010). In the process of constructing trait databases, some entries (trait-by-species combinations) may be missing because no records for a trait are available for a species, and this raises issues with inclusion/exclusion rules for species in an analysis (Nakagawa and Freckleton, 2008). This problem is compounded when trait databases include many traits or poorly sampled species. These missing data are problematic for trait-based approaches in the same way that missing covariates are generally problematic in statistical modeling (Little and Rubin, 2014).

One solution to this missing data problem lies in interpolating missing trait values. Information on missing traits can be derived from closely related species if traits show phylogenetic autocorrelation - that is, if closely related species tend to have similar trait values (Felsenstein, 1985). While phylogenetic non-independence can be a nuisance for many comparative analyses because it violates common assumptions of conditional independence, in the context of trait interpolation it is an asset. Indeed, existing trait interpolation methods such as phylogenetic eigenvector mapping (PEM) make use of phylogenetic autocorrelation to acquire informed estimates of species traits, thereby helping to overcome potential biases in species inclusion (Guénard et al., 2013). Information can also come from other correlated traits recorded from the same species (Westoby et al., 2002; Poorter et al., 2006; Zheng et al., 2009). For instance, large-bodied mammals tend to reproduce more slowly; thus, for a newly described large mammal with no recorded reproduction data, we would expect a long gestation period (Promislow and Harvey, 1990). Our confidence in this prediction will depend on the strength of correlation between the traits. Prediction uncertainty is inherent to interpolation, and ideally this uncertainty would be propagated forward through an analysis that makes use of interpolated traits.

Here, we present a hierarchical Bayesian method for missing-trait interpolation that makes use of latent, life history factors that are structured phylogenetically. This method exploits information from closely related species and from correlated traits to acquire the full posterior distributions for missing traits, conditional on the observed traits. The acquisition of a posterior distribution for missing traits facilitates error propagation, incorporating uncertainty in the model parameters in trait interpolations. This approach simultaneously uses both incomplete and redundant traits with no requirement that any trait be observed for every species (Guénard et al., 2013), thereby expanding its utility to a wide range of situations. Using trait data from 168 primate species, we illustrate the efficacy of this model in predicting unobserved trait values. We conclude with a comparison to existing trait imputation methods, and suggest some future applications and extensions to improve the utility of these approaches.

5.2 Methods

5.2.1 Data description

We compiled primate species trait data from multiple sources, including the Handbook of the Mammals of the World, the PanTHERIA database, and a recently published dataset of species traits for amniotes (Myhrvold et al., 2015). We obtained a phylogenetic tree of 346 primate species from Perelman et al. (2011), and of these, 168 species were represented in the trait data. After sub-setting the trait data to species in the phylogeny, all species traits were logarithm(+1) transformed, centered, and scaled to have mean zero and unit standard deviation. The resulting dataset was a trait matrix with 168 species (columns) and 22 traits (rows), available in the data supplement. In this matrix, 36.4% of trait observations were missing overall (Figure 5.1). No single trait was completely observed, with the percentage of missing data for traits ranging from 0.6% to 77.8%. Further, there were no species with complete trait observations, with the percentage of missing observations at the species level ranging from 9.1% to 95.4%.

5.2.2 Model structure

The aim of our modeling approach was to predict missing species-trait data using incomplete information for a species, as well as any information from closely related species. To achieve this goal, we developed a model that is similar to factor analytic models, where observed data are treated as random variables that are functions of latent factors. In turn, each of these latent factors has some loading coefficient determining the effect on the observed variable (Lopes and West, 2004). For traits $i = 1, \dots, N$ and species $j = 1, \dots, J$ any particular trait observation is z_{ij} , making up the elements of Z which represents the trait matrix. Define y_{ij} to be the transformed trait value, which has undergone a $\log(+1)$ transformation, followed by centering and scaling for each trait. These traits make up the elements in an N by J matrix Y , which contains the transformed trait values. If there are M latent factors, where $M < N$ then each species j is associated with a vector $f_{\cdot j} = (f_{1j}, f_{2j}, \dots, f_{Mj})^T$ of species-specific latent factor values, where the dot subscript in $f_{\cdot j}$ indicates all of the elements across the first index.

To make use of phylogenetic information, we assumed that these latent factors covaried among species, and borrowed ideas developed for spatially autocorrelated factor analysis to allow for phylogenetic autocorrelation (Hogan and Tchernis, 2004). We assumed that the latent factors would be correlated among species, but uncorrelated among each other. We defined a vector of latent factor values for each species $j = 1, \dots, J$, corresponding to each of the M latent dimensions, and imposed a mean zero Gaussian process prior on the vector such that $f_m \sim GP(0, K(d))$, where $K(d)$ is a decreasing covariance function of cophenetic distance d , the pairwise distance between species in the phylogenetic tree, so that closely related species may have more similar latent factor values than distantly related species (Neal, 1998). This implies a multivariate normal prior on all of the latent factors concatenated into vector: $vec(f) = (f_1^T, f_2^T, \dots, f_M^T)^T$, and $vec(f) \sim N(0, \Sigma)$, where Σ is a block diagonal covariance matrix with M blocks, each being a covariance matrix with potentially varying parameters obtained by applying the correlation function $K(d)$ to an N by N square matrix D that contains the cophenetic distances among species obtained from the

phylogenetic tree. Fixing the variance of each latent factor to equal one - a common identifiability constraint - the covariance function effectively becomes a correlation function (Shapiro, 1985).

We considered five correlation functions that varied in their assumptions about the correlations in latent factors as a function of phylogenetic distance (Figure 5.2). Some of these functions such as the exponential allow for long range phylogenetic dependence, while others such as the spherical limit long range dependence. The absence of phylogenetic autocorrelation is represented with a phylogenetically independent correlation function $K_{indep}(d) = I_{d=0}$, which returns an N by N identity matrix for any cophenetic distance matrix D . Theory suggests that multiple phylogenetic correlation functions can result from different mechanistic models of evolution (Hansen and Martins, 1996), but here the goal is simply to interpolate traits, not to represent explicit models of evolution.

Traits arise from the latent factors with process and measurement error. To map the N measured traits to the M latent factors, we use a rectangular loading matrix $\Lambda_{N \times M}$. The upper triangular elements of Λ are fixed at zero, and the diagonal elements $(\lambda_{11}, \lambda_{22}, \dots, \lambda_{DD})$ are positive, ordered, and decreasing (West, 2003; Conti et al., 2014). We used a lognormal prior on the diagonal elements to avoid diagonal loadings of zero, with mean zero and unit variance, and a normal prior on the remaining elements: $\lambda_{im} \sim N(0, \sigma_\lambda) : i \neq m$. Here σ_λ is a hyperparameter that controls the variation in non-diagonal loadings, and it received a half-Cauchy prior: $\sigma_\lambda \sim Cauchy_+(0, 5)$. A matrix $F_{M \times J}$ contains the latent factor values such that element $F_{m,j}$ corresponds to latent dimension m and species j . The N by J matrix product ΛF provides the mapping of latent factors to the observations. We also included parameters for the means $\alpha_1, \dots, \alpha_N$ and variances $\sigma_1^2, \dots, \sigma_N^2$ for each trait, although we expected these to be close to zero and one respectively after centering and scaling the data. Therefore, we used informative normal priors centered around zero for the means $\alpha_i \sim N(0, 0.1)$, and informative half-Cauchy priors for the standard deviations $\sigma_i \sim Cauchy_+(1, 0.1)$. For notational convenience, we also define a matrix $A_{N \times J}$, with the $(i, j)^{th}$ element $\alpha_{ij} = \alpha_i$. Then the model can be written compactly as $Y = \Lambda F + A + \epsilon$, with $\epsilon_{N \times J}$ being a matrix of normal residuals with trait-specific distributions: $\epsilon_{ij} \sim N(0, \sigma_i)$.

We assumed that traits were missing completely at random - that is, the presence or absence of an observation was independent of its value (Little and Rubin, 2014). This assumption may be violated if for instance, smaller species are harder to measure, such that missing body mass observations are a non-random sample of the body mass distribution. If there were known missingness mechanisms, they could be modeled, with the potential to increase the precision with which missing traits are predicted (Reich and Bandyopadhyay, 2010).

5.2.3 Parameter estimation

We used automatic differentiation variational inference, implemented via the Stan probabilistic programming language, to approximate the posterior distribution of the parameters (Hoffman and Gelman, 2014; Kucukelbir et al., 2015) . This algorithm is considerably faster than traditional Markov chain Monte Carlo methods, allowing us to estimate parameters for many models with a relatively large dataset in ≈ 48 hours on a 4-core laptop.

5.2.4 Evaluating predictive performance

We used K-fold cross validation to compare the predictive power of the models in our model set, which included models with the five different latent factor correlation functions, and up to 17 latent factor dimensions (Arlot and Celisse, 2010). Each observation was assigned to one of $K = 10$ folds at random, generating K test data sets y_k , with $k = 1, \dots, K$. For any fold, the observations not assigned to that fold comprise a training data set, y_{-k} , which is used to obtain a posterior distribution for the parameters $p(\theta | y_{-k})$. Predictive performance for observation $i \in k$ is quantified as the log predictive density, averaging over S posterior simulation draws: $\widehat{lpd}_i = \log(S^{-1} \sum_{s=1}^S p(y_i | \theta_{k,s}))$ (Gelman et al., 2014). If \widehat{lpd}_i is high, then y_i is predicted well by the model. For each model, we obtained a summary measure of predictive performance by summing over N to obtain the expected log predictive density $\widehat{elpd} = \sum_{i=1}^N \widehat{lpd}_i$ (Vehtari et al., 2015). To calculate uncertainty in model performance, we also calculated the standard error of the expected log predictive density $se(\widehat{elpd}) = \sqrt{N \text{var}(\widehat{elpd})}$, where var is the sample variance function

computed over all observations $i = 1, \dots, N$.

5.3 Results

In predicting unobserved traits among the included primate species, the model with a squared exponential correlation function and seven latent factors exhibited the highest predictive performance (Figure 5.3). For this model, posterior trait expected values correlated strongly with the held-out test values with the median correlation coefficient $r_{med} = 0.856$, and 95% posterior credible interval (CI): (0.802, 0.905). In addition, the posterior predicted trait values - which include both inferential and predictive error - also were well-correlated with the held-out values, with $r_{med} = 0.793$ and CI: (0.675, 0.854) (Figure 5.4). Interval coverage was as expected with 95% of the 95% posterior prediction intervals containing the true trait values (Figure 5.5). While the squared exponential, seven-factor model performed best, other models with exponential and rational quadratic correlation functions were also well supported, whereas both the phylogenetically independent and spherical correlation function models performed poorly (Figure 5.3).

Predictive performance was best for traits that were well sampled, while traits with few data points had higher predictive error (Figure 5.6). The most poorly sampled traits, such as the lower and upper bounds on body height, were predicted rather poorly, and should be taken with skepticism (Figure 5.7). Predictions for these traits varied little, being closely centered around the mean values, and also showed low precision, with wide posterior credible intervals, unlike predictions for the well-sampled traits (Figure 5.7).

5.4 Discussion

By taking advantage of incomplete trait records and phylogenetic autocorrelation, the latent-factor approach developed here facilitated estimation of missing species traits and performed well based on cross validation. The most well-supported model made use of a squared exponential phylogenetic correlation function for the latent factors, for which closely related species are very similar, and distantly related species are dissimilar. In contrast, the model with no phylogenetic

structure showed poor predictive performance (Figure 5.3).

In addition to acquiring estimates of missing species traits, this method also offers insights about which latent factors are correlated among species, but uncorrelated across the M latent dimensions. These latent factors could be used as explanatory variables for other species-level characteristics, including estimates of extinction risk, and may outperform species traits in this regard because they are uncorrelated with each other, resulting in less variance inflation for coefficient estimates (Stine, 1995). In particular, this approach could prove useful for data-deficient species with well-described phylogenetic neighbors that provide information on the poorly sampled species' latent factor values (Morais et al., 2013).

This Bayesian latent factor approach has some potential advantages and disadvantages compared to existing phylogenetic trait interpolation methods. Compared to phylogenetic eigenvector mapping, which uses phylogenetic topology and branch lengths to predict missing species traits one at a time, this method operates on all traits simultaneously (Guénard et al., 2013). To improve predictions, fully observed traits can be added as covariates in PEM approaches, but this is limiting because (as in our case) there may be no traits with records for each species. One solution would be to subset the data so that at least one trait was fully observed; however, this type of omission runs counter to the spirit of trait interpolation and unnecessarily discards information. Further, as previously noted, datasets for traits are often so incomplete that creating a subset of any large group of organisms to complete a trait may make for such a small sample that analysis becomes non-informative for the larger group. The latent-factor approach described here does not require that any particular trait be completely or even mostly observed, and also is inherently multivariate, making use of correlations among incompletely observed traits to make predictions for all missing records simultaneously. However, unlike PEMs, the latent-factor approach does not make use of an explicit mechanistic model of evolution to predict missing species traits (Alexandre Felizola Diniz-Filho, 2001; Felizola Diniz Filho et al., 2012). If the primary goal is to evaluate support for different models of trait evolution *and* to predict missing species traits, PEMs may be preferred over latent factor approaches.

As trait-based approaches become more common and more trait databases are assembled, demand for trait interpolations will increase. This latent factor method performs well and is very flexible, making use of any number of incompletely observed traits while exploiting phylogenetic autocorrelation and among-trait correlations to make better predictions. In addition, this model can easily be embedded within other hierarchical models, and make use of other trait distributions (e.g., discrete valued traits modeled with binomial or Poisson likelihoods). However, both the latent factor approach and the PEM approach are currently limited in that they require complete phylogenetic information for all included species. Incorporation of phylogenetic uncertainty is thus an important future direction for both methods, especially if species with missing trait data also tend to be missing from phylogenetic trees.

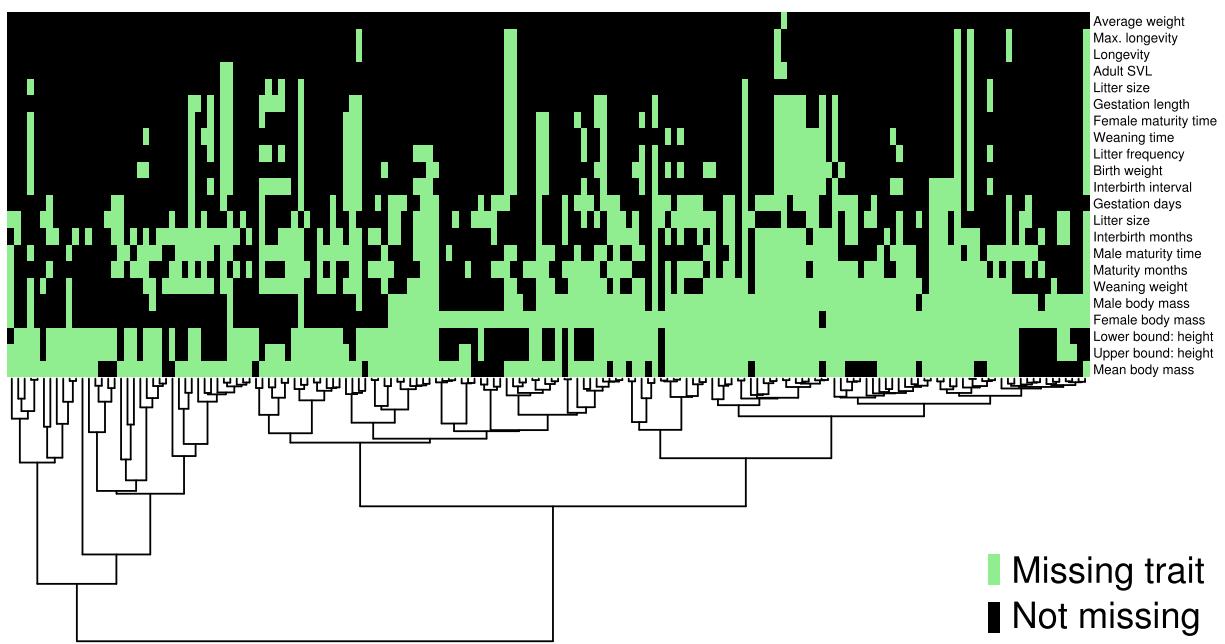


Figure 5.1: Missing trait map matched to the species level phylogeny. The matrix in the upper half represents the trait data, with missing values shown in green. The columns of the matrix represent species, for which phylogenetic relationships are illustrated below. Rows that are mostly green represent poorly sampled traits, and columns that are mostly green represent poorly sampled species.

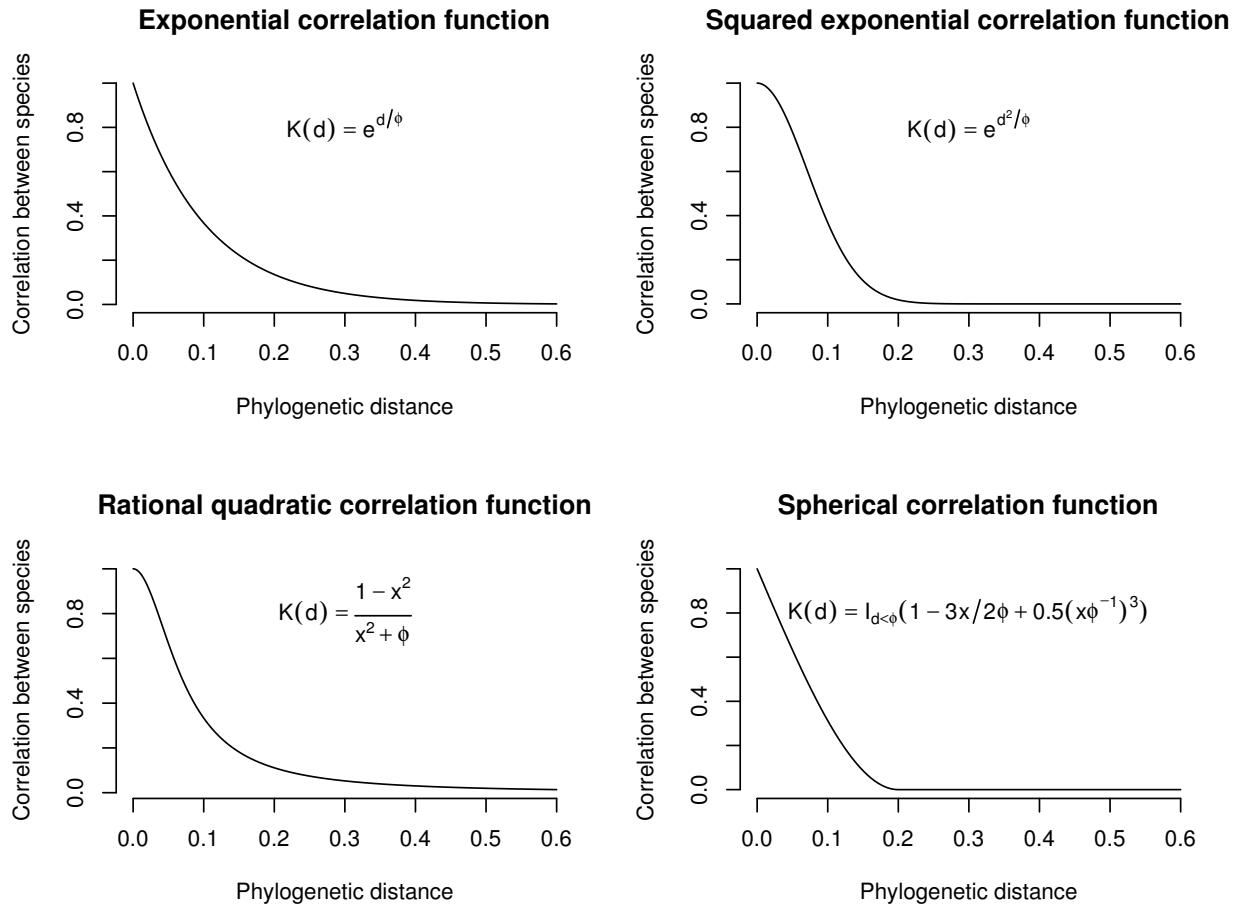


Figure 5.2: Correlation functions considered for the latent factors. The x-axis represents cophenetic distance between species, and the y axis represents the correlation induced by the function. Each function has one parameter ϕ , but the shapes of the functions differ.

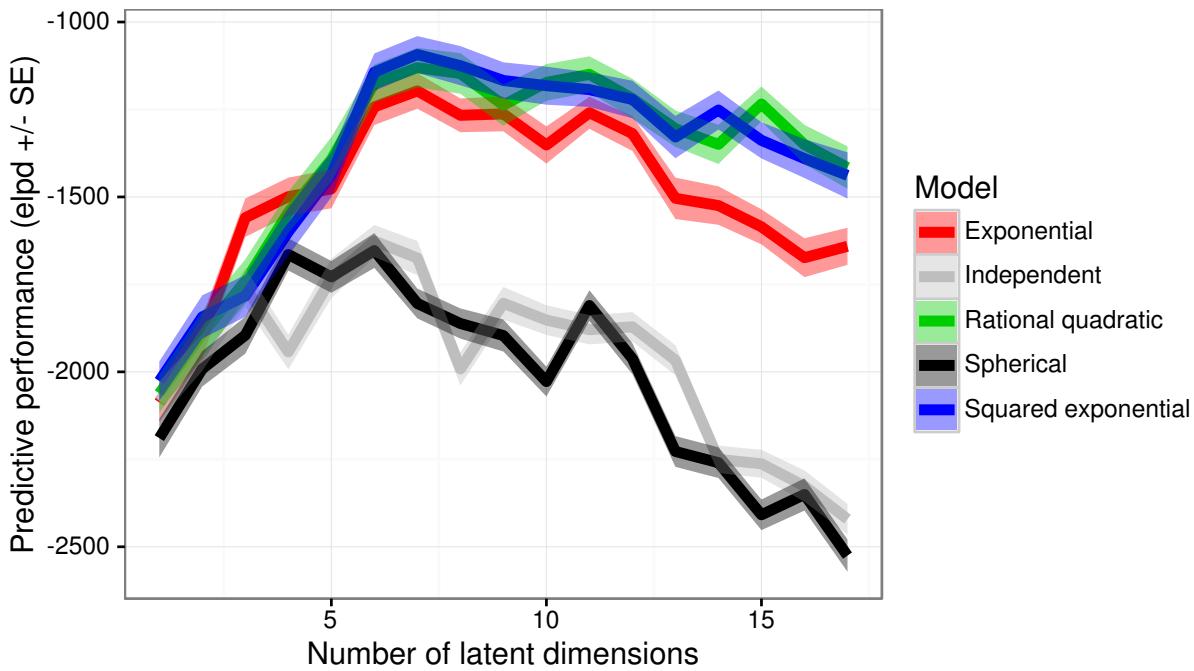


Figure 5.3: Performance curves for models with varying phylogenetic correlation structures (indicated by color), and latent factor dimensionality (shown in the x-axis). The y-axis is a measure of predictive performance for the model across all 10 folds. Specifically we show the expected log predictive density (elpd) plus and minus the standard error. High values of elpd indicate that the held out test data were consistent with the model predictions generated with the training data, and low values indicate that a model's predictions did not match the test data well.

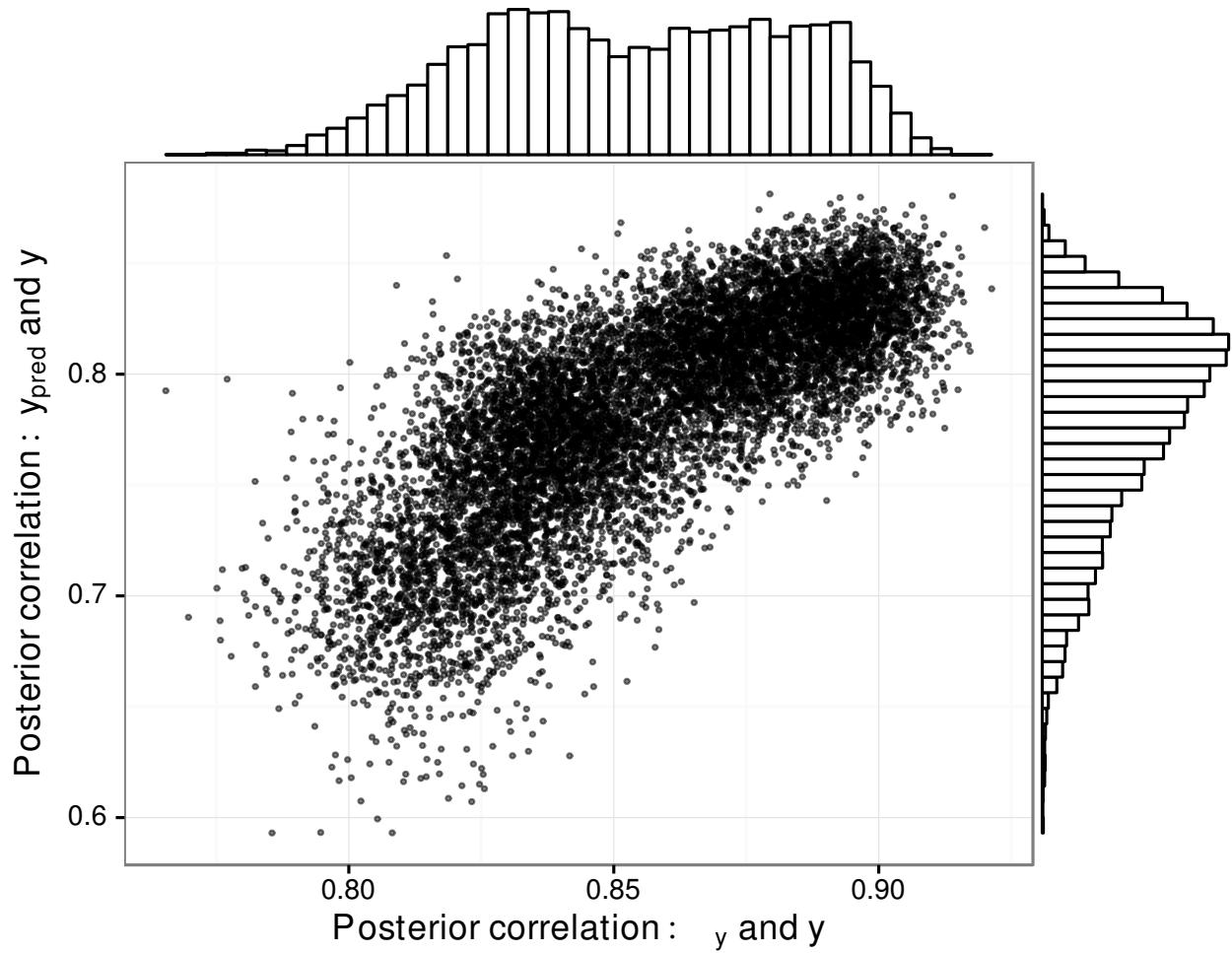


Figure 5.4: Bivariate posterior distribution for the correlation between held-out values and posterior expected values (x-axis), and between the held-out values and the posterior predicted values (y-axis). Each point is a draw from the posterior distribution, and all draws are shown for each of the 10 folds. Marginal posterior histograms are shown for both quantities.

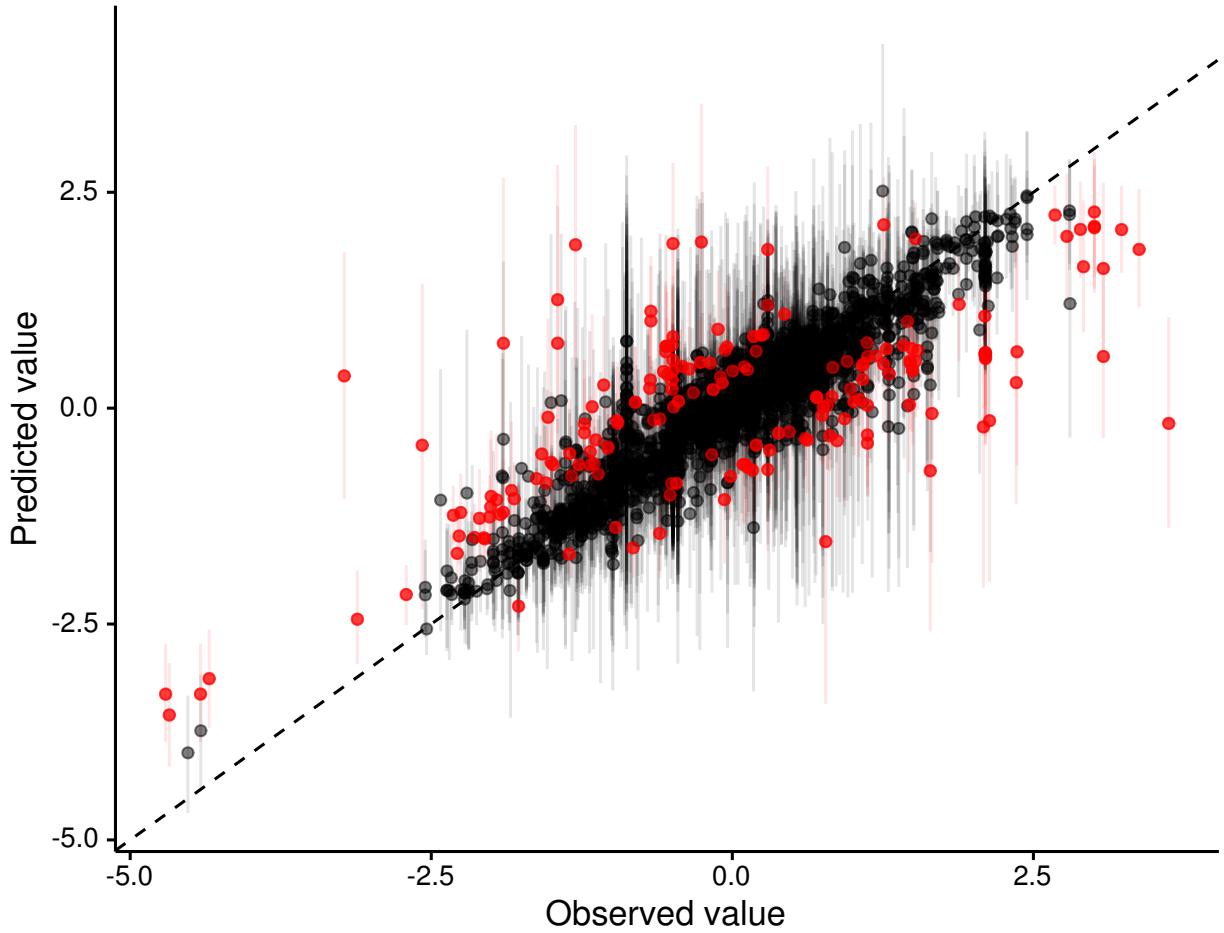


Figure 5.5: Observed (held out) values and model predictions. Each point is a species trait, with the x position representing the true (scaled) value, and the y position of the point represents the posterior median prediction from a model with seven latent factors and a squared exponential phylogenetic correlation function. The vertical line segments represent 95% credible intervals for the predicted trait values. The dashed line is a one to one line. If a point lies on this line, then the prediction was centered around the true value. Red indicates that the true held-out trait value was not in the 95% prediction interval, and the red points have been plotted over the black points to highlight model misfit.

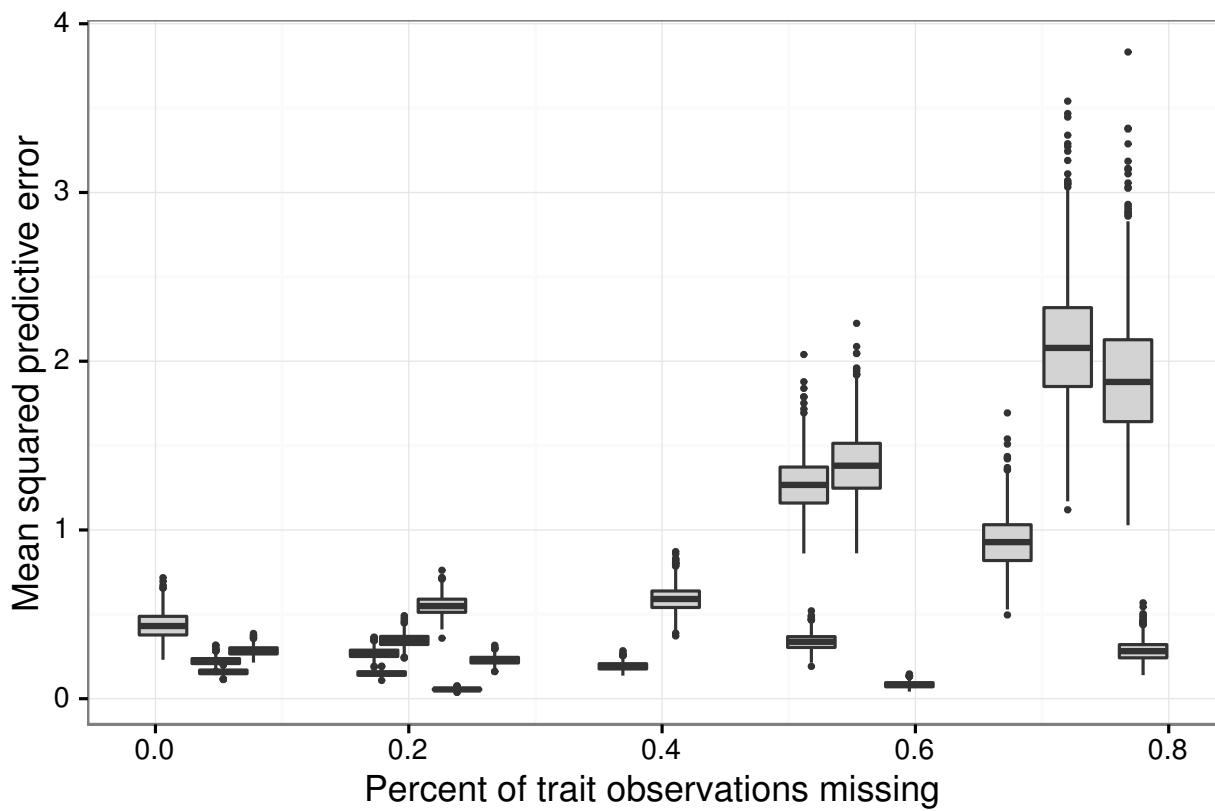


Figure 5.6: Mean squared predictive error as a function of the number of trait observations. Each boxplot represents one trait. Well sampled traits are on the left half of the graph, and poorly sampled traits are on the right half. The midpoint in the box represents the posterior median, and the posterior quantiles are shown as the upper and lower portions of the box. Whiskers

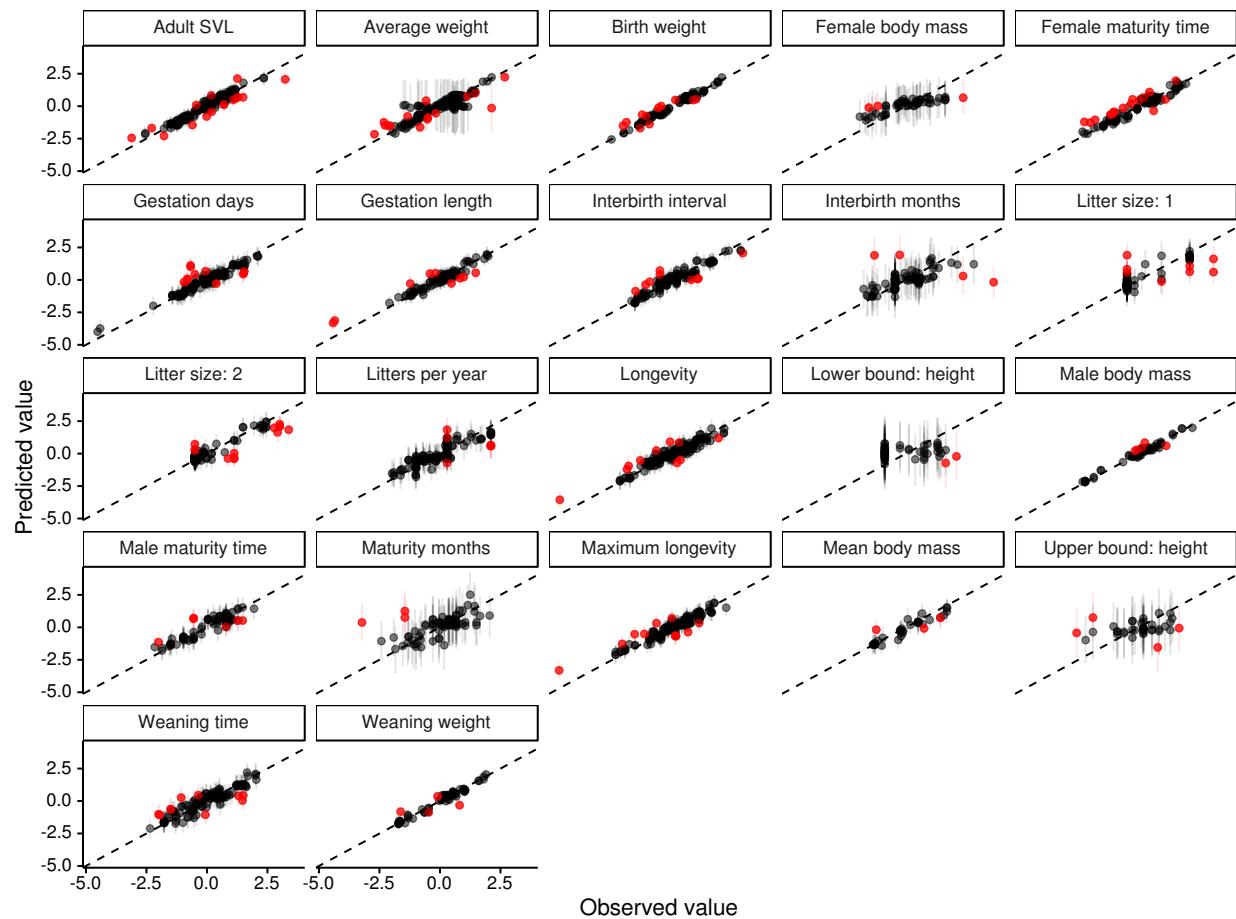


Figure 5.7: Posterior predictions with 95% prediction intervals, separated by trait. True values are on the x-axis, and the y-axis location of each point is the posterior median of the predicted value. Red indicates that the true held-out trait value was not in the 95% prediction interval, and the red points have been plotted over the black points to highlight model misfit.

Chapter 6

Host diversity begets symbiont diversity but reduces symbiont transmission

MAXWELL B. JOSEPH, JOSEPH R. MIHALJEVIC, PIETER T. J. JOHNSON

Explaining patterns of biodiversity is one of the primary aims of ecology, and historically these patterns have been described in terms of macroorganisms such as plants and animals. Recent methodological advances grant the ability to observe previously unobservable microbial communities, which are incredibly diverse and often nested within free-living organisms, providing an impetus and means to expand understanding of the mechanisms that structure communities. These mechanisms are potentially similar to the mechanisms that structure free-living communities, with habitat patches for free-living species being analogous to host individuals for symbiotic organisms, often microbial, that live in or on free-living species. Here, we apply this analogy to explore the possible effects of host functional diversity on symbiotic organisms with a stochastic multi-species model across a range of scenarios including specialist, generalist, parasitic, commensal, and mutualistic symbionts. The model predicts that host diversity nearly always increases symbiont diversity, though the shape of the increase varies as a function of symbiont niche breadth and the fitness consequences of infection. In addition, diverse host communities are predicted to have lower transmission rates on average, and relatively greater transmission of generalist symbionts. The same mechanisms allow host diversity to increase symbiont diversity while reducing transmission. Namely, diverse host communities have higher symbiont colonization rates but fewer hosts available for local transmission, which increases the probability of stochastic symbiont extirpation. These results are potentially useful as a step toward a more unified community ecology of free-living and

symbiotic organisms.

6.1 Introduction

Recent progress in disease ecology has highlighted the importance of host diversity for infectious disease dynamics, but relatively little is known about the role of host diversity in regulating the dynamics of transmissible symbionts in general (Lively et al., 2014; Johnson et al., 2015). Specifically, the joint influence of host diversity on symbiont transmission and diversity remains poorly understood. Theory has been developed to understand how host community composition affects the transmission dynamics of symbiotic pathogens (Dobson, 2004; Roche et al., 2012; Joseph et al., 2013; Mihaljevic et al., 2014), and biogeographic studies indicate positive correlations between host and parasite diversity (Nunn et al., 2005; Dunn et al., 2010; Harris and Dunn, 2010). Methodological advances may now facilitate joint consideration of these phenomena, as next generation sequencing technologies increasingly reveal community composition of symbionts within host populations and communities (Prosser et al., 2007; Fierer et al., 2012; Costello et al., 2012). Despite this opportunity to observe symbiont communities, we currently lack theoretical predictions about host community composition influence on symbiont diversity and transmission dynamics, possibly due to computational challenges in modeling these complex systems across multiple scales (Gog et al., 2014). Here, we make an effort to overcome these challenges, presenting a general model that explores how host biodiversity affects the transmission and biodiversity of parasites, mutualists, and commensal symbionts.

To predict how host diversity generally influences symbiont diversity and transmission, we may be able to apply ecological and biogeographical concepts developed for free-living communities to symbiont communities inhabiting multiple host species. For instance, analogies have been drawn between free-living communities living within habitat patches and symbiont communities living inside host individuals (Mihaljevic, 2012; Richgels et al., 2013; Suzán et al., 2015). One way to view host biodiversity from the symbionts' perspective is through the concept of habitat heterogeneity, where individuals from different host species present unique environmental conditions that could

influence the population dynamics of symbionts. Given this analogy, we can then test whether predictions from free-living systems hold for symbiont communities, which have unique dynamics compared to free-living species.

In free-living systems habitat heterogeneity can increase species coexistence locally and regionally, because diverse habitats contain resources that can be exploited by multiple species and source-sink dynamics are more likely to rescue species from local extirpation (Levins, 1969; Pacala and Roughgarden, 1982; Amarasekare et al., 2004). For a habitat region of fixed area, as habitat heterogeneity increases, the area of each habitat type must go down because of an inherent trade-off with habitat heterogeneity. As a result, mean population sizes have been predicted to decrease with habitat heterogeneity, leading to higher probabilities of stochastic extinctions (Allouche et al., 2012). Thus, when habitat heterogeneity increases, but habitat area remains fixed, a unimodal relationship between heterogeneity and free-living species richness is expected due to low persistence probabilities at very high habitat heterogeneity (Allouche et al., 2012). However, it remains an open question whether this ecological theory pertaining to habitat heterogeneity is relevant to explaining patterns of symbiont diversity and transmission.

In this study, we use a computational model to explore whether host diversity influences symbiont communities in a way analogous to the effects of habitat heterogeneity in free-living systems. We do this by simulating communities of many host and symbiont species and manipulating host functional diversity and characteristics of the symbiont community. Our model is unique in that it incorporates ecological dynamics important for free-living host population dynamics, such as colonization and reproduction, as well as mechanisms exclusive to host-symbiont interactions, such as transmission and symbiont-mediated fitness effects on the host. We use this modeling framework to determine how host functional diversity could affect symbiont transmission and symbiont species richness over a range of scenarios. Specifically, in our model we can manipulate symbiont niche breadths - allowing for a combination of specialist and generalist symbionts - and symbiont-mediated fitness effects - allowing for a combination of parasites, commensals, and mutualists. This allows simultaneous investigation of the effect of host biodiversity on local symbiont transmission

and diversity.

6.2 Methods

We developed a stochastic, continuous-time, agent-based model for the dynamics of host-symbiont communities. In the model, a host community resides in a habitat patch that is divided into a finite number of cells C , i.e. micro-sites which can be occupied by a single host individual. Within the habitat patch, host colonization, reproduction, and death occur with rates c , r , and d , respectively, which are constant across model realizations. Multiple host species $h = 1, \dots, H$ exist within a regional pool, and the patch experiences a constant rate of host colonization c , which is equal for all host species, and there is no displacement of occupying hosts by colonizing hosts. Host species vary in traits that are relevant to symbiont infection, but the communities are otherwise neutral, in that all host species have equal colonization, reproduction, and death rates. Once a host colonizes a cell, it then has some rate of reproduction and of death that is constant among host species. Offspring attempt to disperse to a random habitat patch cell, and if it is unoccupied, they successfully colonize.

Multiple types (species) of symbionts occur in the regional pool and attempt to colonize host individuals with a constant rate r_s . Host species vary in traits relevant to symbiont establishment probability. We assume for simplicity that this variation is unidimensional, as would be the case if there were one primary niche axis for the symbionts, e.g. pH of the host gut. Symbiont niches are represented as univariate Gaussian functions with niche optima and breadths (means and variances, respectively), such that the probability of successful establishment p_e on an individual host is a function of the host's value along the niche axis (Figure 6.1). Symbionts have the potential to establish when a susceptible host contacts an infected host, or when a symbiont from the regional pool attempts to colonize a susceptible host in the habitat patch.

Every individual host has some constant probability of a symbiont from the regional pool attempting to infect, with an infection establishment probability p_e from the symbiont niche function. If a symbiont successfully colonizes a host, it becomes part of the local community and can either be

extirpated through host recovery or persist by transmitting to other hosts. Every infected host has a recovery rate that is assumed constant across all species. Hosts contact each other with rate ϕ , also assumed constant across host species, so that the overall rate of contacts between susceptible and infected individuals is ϕSI , where S is the number of susceptible hosts, and I is the number of infected hosts. Conditional on a contact, the infected host passes the symbiont infection to the susceptible host probability p_e , derived from the symbiont niche function. In this way, host species that are functionally similar are more likely to exchange symbionts.

We implemented this model via stochastic simulation, generating continuous-time Markov chains in high-dimensional space, where the dimensions represent the states: number of hosts of each species, and the symbionts that infect each individual. Although continuous-time Markov processes are more difficult computationally than discrete-time formulations, they have the advantage of not requiring an arbitrarily specified order of events at each time step. The order of events emerges from the rates of each potential process via the Gillespie algorithm (Gillespie, 1976). Given the state of the system at any time point, every potential event that could happen is represented by some rate r_e for events $e = 1, \dots, E$. The total rate of events is simply the sum of the component rates $r_{tot} = \sum_1^E r_e$, and the time until an even happens is exponentially distributed, with overall rate parameter r_{tot} . This provides a way to generate the timepoints stochastically, and given some time of an event happening, the specific event is be selected with probability e_r/r_{tot} .

We were interested broadly in the effect of host functional diversity on symbiont transmission and richness within a local host community. To vary host functional diversity, we specified a range of within-host environmental conditions and uniformly simulated values bounded within this range for each host species, and assigned these values to host species in the regional pool. The realized functional diversity within a local community is the range of values represented by host species that occupy the local community, and can thus vary across a simulated time series as new host species colonize or are extirpated. We monitored the number of symbiont transmission events and symbiont species richness throughout each simulated time series. In order to compare values across many simulated time series, we restrict our calculations of mean host functional diversity, symbiont

transmission, etc. to common periods in which the quantities have stabilized, rather than including the entire time series. While the contact rate ϕ is a model parameter, the transmission rate for any time period is a model outcome, calculated as the number of transmission events divided by the amount of time.

In addition to investigating the effect of host functional diversity on symbiont transmission and richness, we explored the effect of symbiont niche breadth (σ in Figure 6.1), and the effect of infection on host fitness, described below. These questions lead to six different simulation experiments. The first explores the effect of host functional diversity on transmission and symbiont richness when all symbionts are functionally equivalent and commensal. The second allows symbiont niche width to vary among simulated realizations, but assumes that the niche width of all symbionts is equal within a simulation run. The third varies the effect of symbiont infection on host survival, varying across simulation runs but equal among symbionts within a run. The fourth varies symbiont niche width within and across simulation runs so that the regional pool of symbionts includes specialists and generalists. The fifth varies symbiont fitness effects within and across runs so that the regional pool includes parasites and mutualists. The sixth and final experiment allows both niche width and fitness effects to vary within and across runs so that the regional pool of symbionts includes a range of specialist and generalist parasites and mutualists. All code to replicate the analysis and manuscript is publicly available at <https://github.com/mbjoseph/abm>.

6.3 Results

In our first experiment all symbionts were commensal and had equal niche widths. Across 1000 simulated local host communities, host functional diversity increased symbiont richness and decreased transmission (Figure 6.2). Symbiont diversity increased non-linearly with host functional diversity, with the fastest increases occurring at low levels of host functional diversity. While symbiont transmission was lower on average in communities with high symbiont diversity and high host functional diversity, there was one instance in a low-diversity host community in which no symbionts could colonize, because the range of host resources that were available in the local

community was outside the optimal niche space of the symbionts in the regional pool. In other words, symbionts attempting to colonize had very low probabilities of establishment in the hosts, such that across the range of timesteps, no symbionts succeeded in colonizing the local community.

In our second experiment, where the niche width of symbionts varied among simulations but was constant among species within each simulation, host diversity still increased symbiont richness and decreased transmission, and these relationships were strongly moderated by symbiont niche width (Figure 6.3). The increases in symbiont richness with host richness were still non-linear, but they shifted from being concave down to concave up as symbionts became more generalized. Transmission consistently decreased as host communities became more functionally diverse, but as symbiont niche width increased, there were fewer outlier communities with extremely low transmission, because it becomes less likely that any particular host community will have a range of resources available that cannot be exploited by symbionts in the regional pool because of a mismatch with their niche space.

In our third experiment, where the fitness consequences of infection varied among simulations, host diversity still increased symbiont richness and reduced transmission, but parasites had lower richness and transmission on average compared to mutualist and commensal symbionts (Figure 6.4). Parasites restrict their own transmission by killing hosts, making it more difficult to persist in a local host community (Wood and Thomas, 1996). As a result, richer host communities still can be colonized by more parasitic symbiont species, but these communities will have more depauperate symbiont communities relative to communities comprised of less virulent symbionts.

When symbiont niche width varied within communities, host diversity still increased symbiont richness, but the effects of host diversity on symbiont transmission were more subtle (Figure 6.5). Specialist symbionts had the highest transmission rates in low-diversity communities, but had relatively low transmission in more diverse host communities. Generalist symbionts had higher transmission in the high diversity host communities, and lower transmission in less diverse host communities, consistent with the idea that specialist symbionts can outcompete generalists. In addition, mean niche breadth in the symbiont community increased with host functional diversity,

such that diverse host communities contained more generalist symbionts than low diversity host communities.

When symbionts varied in fitness effects, host diversity increased symbiont richness, and reduced transmission, but the reduction in transmission was more pronounced for parasitic than mutualistic symbionts (Figure 6.6). Interestingly, extremely mutualistic symbionts had a lower peak transmission in low diversity communities than moderately mutualistic symbionts, because hosts were so long-lived that the majority of habitat patches in the lowest diversity communities were occupied by extremely long-lived, infected hosts, such that there were fewer susceptible hosts available for transmission.

When symbionts varied in fitness effects and niche breadth, host diversity still increased symbiont diversity and reduced transmission (Figure 6.7). Mutualistic symbionts nearly always had higher transmission than parasitic symbionts, and on average, symbionts in the local communities were mutualistic. Furthermore, low diversity host communities had more specialist symbionts than high diversity communities, which primarily consisted of generalists. Both of these findings are consistent with the component cases when fitness effects and niche breadth are the only varying parameters.

6.4 Discussion

Across a wide range of situations, increasing host functional diversity consistently increased symbiont diversity and reduced average transmission in our model. Fitness consequences of infection and symbiont niche breadth regulated the degree and, to a lesser extent, the shape of these relationships. In general, we expect that across replicated communities in nature, with all else being equal, host communities that are low in functional diversity will have relatively low symbiont richness but high average transmission per symbiont species, and that specialist symbionts may be more likely to occur in high abundance. In contrast, host communities that are highly diverse are more likely to have diverse symbiont communities with lower mean transmission rates, and also more generalist symbionts.

Recent theory has clarified conditions under which host biodiversity might affect transmission of parasitic symbionts, but relatively few studies have considered multiple symbiont species simultaneously. For instance, we know that host community richness, functional diversity, evenness and the distribution of life-history traits can all affect the transmission of a focal pathogen (Dobson, 2004; Rudolf and Antonovics, 2005; Roche et al., 2012; Joseph et al., 2013; Mihaljevic et al., 2014; Chen and Zhou, 2015; ORegan et al., 2015). Here we have generalized previous theory to include multi-host, multi-symbiont dynamics, allowing us to consider the effects of host biodiversity on patterns of symbiont diversity and transmission simultaneously. Further, we have expanded previous models to allow for symbionts with variable effects on host fitness, including parasitic, mutualistic and commensal organisms, which more closely matches the composition of natural symbiont communities. Under a broad set of assumptions, host diversity promoted more diverse symbiont communities, and symbiont diversity scaled linearly, concave up, and concave down with host diversity. All of these functional forms have been predicted by analytical models for parasites in host communities undergoing disassembly and have been observed in some host-parasite food webs Lafferty (2012). In addition, consistent with previous models that vary host diversity and in which host density does not systematically increase host diversity (Mihaljevic et al., 2014), we found that transmission tends to be reduced in diverse host communities (Dobson, 2004; Rudolf and Antonovics, 2005; Roche et al., 2012; Joseph et al., 2013; Mihaljevic et al., 2014).

Here we have shown that these general predictions may hold not only for parasites, but also for commensal and perhaps mutualistic organisms, and that the same mechanisms underlie both patterns of increased richness and reduced transmission. Specifically, in low diversity host communities, all hosts are functionally similar so that they readily share symbionts upon contact - leading to high transmission. Yet, the pool of potential symbiont colonists is small, restricted to the subset of symbionts that can effectively exploit the narrow range of host resources that are represented in the local community - leading to low symbiont diversity. In high diversity host communities, different host species generally have unique resources available for symbionts, so that when any two individual hosts of randomly chosen species make contact, they are less likely to transmit

their symbionts. In addition, because we have fixed community carrying capacity, increases in host diversity imply that there are fewer individuals of each host species present. Thus, there are fewer host 'habitats' that represent any particular set of conditions for symbionts, resulting in a decrease in transmission. Simultaneously, the number of symbionts in the regional pool that can establish in the local community is greater in diverse communities, where a broader set of resources are represented. So, while more symbiont species may colonize the local community, on average their transmission is reduced. These results have strong connections to theory developed around relationships between habitat heterogeneity and free living diversity.

Allouche et al. (2012) predicted that for a fixed habitat area, increases in habitat diversity may lead to concave down and even unimodal relationships between habitat heterogeneity and free-living richness, and the mechanisms underlying this pattern are analogous to the mechanisms discussed in this paper. Hosts have long been compared to habitat patches for symbionts, except these habitat patches are generally mobile, reproduce, compete with and contact each other, and experience differential survival and reproduction as a result of symbiont occupation or infection (Kuris et al., 1980; Mihaljevic, 2012). In free living species, for fixed area, increasing habitat diversity tends to reduce the mean area of each habitat type. Similarly, for a fixed number of host individuals, increases in host diversity reduces the number of any particular host. Increases in habitat or host diversity increase the colonization rate within local communities, while decreases in the abundance of a habitat or host type increase the probability of stochastic extinctions for free-living and symbiotic species. While we have not included the hosts' habitat heterogeneity in this model, it would be a logical next step to explore how this habitat heterogeneity could jointly influence host and symbiont diversity. This type of modeling effort, however, would require more assumptions about the relationship between host and symbiont niches, which unfortunately are currently ill-informed by empirical data.

As it has been currently formulated, the predictions from this model are broadly consistent with existing empirical data on the apparent effects of host diversity on symbiont transmission and diversity. Specifically, on average diverse host communities have lower abundances of parasitic

symbionts (Civitello et al., 2015), although this claim and topic in general have been contentious (Salkeld et al., 2013; Lafferty, 2012). While parasite abundance is not a direct measure of transmission, theoretical models predict that the reduction of transmission is a likely mechanism underlying the abundance pattern (Dobson, 2004; Rudolf and Antonovics, 2005; Mihaljevic et al., 2014). Here we have broadened the focus to include commensal and mutualistic symbionts, and this model's expectations for the transmission of parasites are not vastly different than the predictions for transmission of mutualists and symbionts. Broadening the focus of the diversity-disease literature to include mutualists and symbionts is more important and feasible now more than ever, given our capacity to collect massive amounts of data on microbes using new sequencing methods. Aside from transmission, it has long been recognized that host diversity facilitates parasite diversity, and this model provides a basis for this relation along with more broadly applicable predictions for specialist vs. generalist and parasitic vs. mutualistic symbionts.

While this model represents a step toward more realistic representations of the dynamics of multi-host, multi-symbiont systems, as with any model, we have made simplifying assumptions that could be relaxed in future studies. Specifically, by including many species of hosts and symbionts, our model has many parameters, and with a simulation approach we can only explore a subset of parameter space. While an analytic model would be useful, the mathematical complexity of such models increases rapidly as species are added. In addition, we assumed that hosts could only be singly infected - that they could not be infected with multiple symbiont species simultaneously. We know that co-infection is common in nature, but including co-infection in the model would require additional sets of assumptions about interactions between symbionts within host individuals (e.g., can symbionts displace each other, is there facilitation among symbionts, competition, neutral interactions, etc.), reducing the tractability of the model. In addition, we also know that local habitat patches are often connected, raising questions about the metacommunity dynamics of the hosts and symbionts, which can be considered to exist within their own metacommunities among host individuals within a community (Mihaljevic, 2012), however in this model we did not explicitly model metacommunity dynamics among habitat patches. Last, we have assumed direct horizontal

transmission for all symbionts, but in nature, many other forms of transmission are common, particularly for mutualistic symbionts which are commonly vertically transmitted.

Beyond these ecological phenomena that were not included, we also did not include evolutionary change in either the symbionts or the hosts. However, this represents a key extension to consider in the future, as the confluence of ecological and evolutionary timescales is one of the hallmarks of microbe-host dynamics. There are many issues to consider from an evolutionary perspective for these types of models, including the evolutionary pressures imposed by the assembly and disassembly of host communities, and the selective effects of habitat or host diversity on symbiont traits. Finally, while this model may be useful in generating qualitative predictions, because it is not parameterized for any particular system or set of species, validation may be more difficult. The predictions from this model may be best considered to be broad expectations for patterns seen across many replicated communities.

Taken together, the predictions for this model serve to unite two previously separate ideas, that diverse host communities may have more diverse symbiont communities and that diverse host communities can lead to lower transmission of a focal symbiont. Our perspective, which includes parasites but also commensals and mutualists, is more general and probably less prone to vitriol than the perspective under scrutiny in the diversity-disease debate, which is focused on the somewhat vaguely defined outcome of disease risk (Ostfeld and LoGiudice, 2003). Possibly, a consideration of symbiont diversity and transmission as an outcome is more precise. There are many additional questions to be explored regarding the relationship between host and symbiont diversity, and we have attempted to outline these above. These theoretical questions have empirical counterparts for future work in this arena which has the potential to leverage new technologies in entire community sequencing to rapidly acquire data.

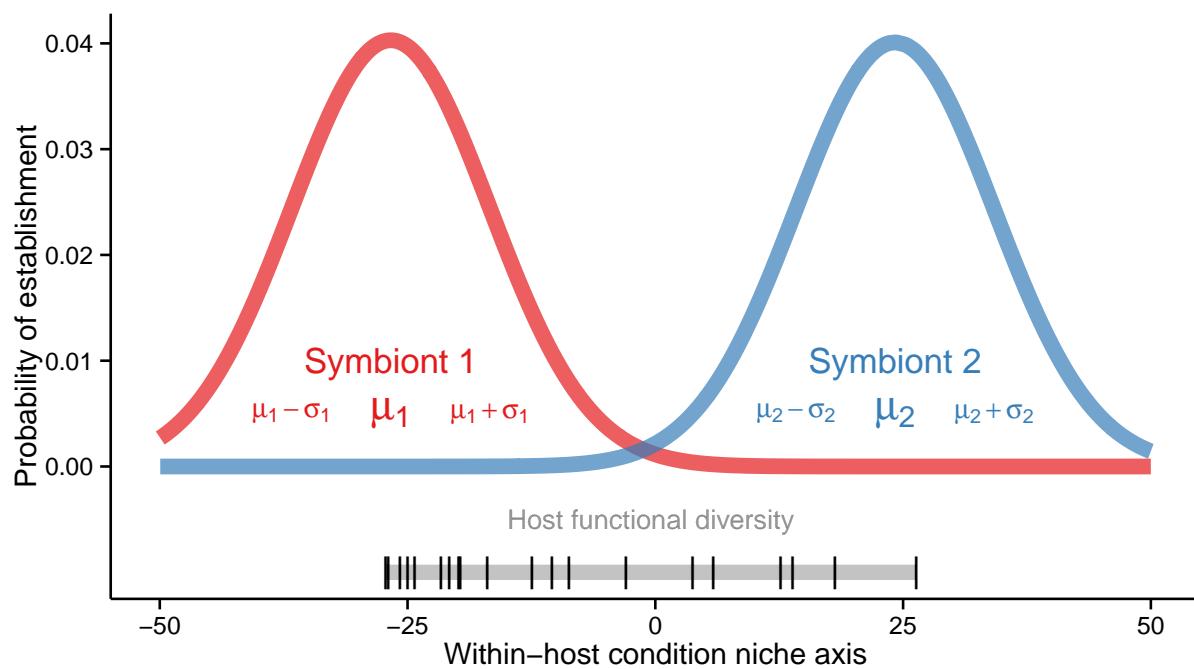


Figure 6.1: Conceptual diagram illustrating the niche space of symbionts along a univariate host-condition axis (the x-axis). Each host species in a local host community is assigned one particular value, shown as vertical bars. The range of the values represents the host functional diversity represented in the local community. Here two symbiont species have equal niche breadths σ but different optima μ , which together define the Gaussian function that relates host condition to the probability of symbiont establishment, conditional on an infection attempt.

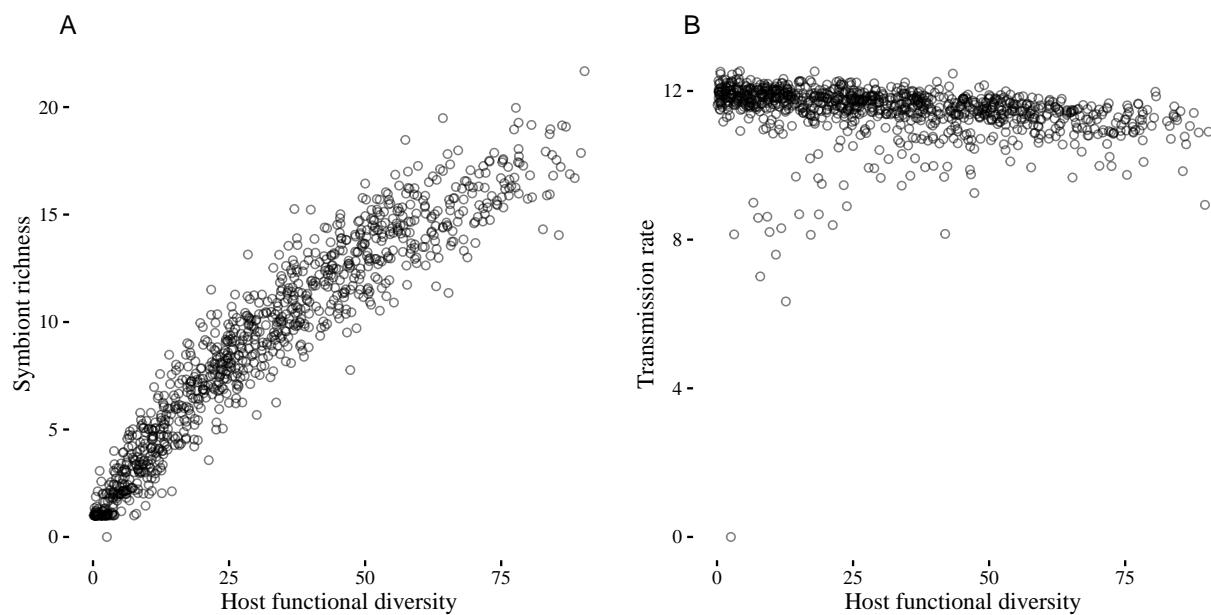


Figure 6.2: Model predictions for commensal symbionts that all have the same niche width. Panel A shows the relationship between mean host functional diversity and symbiont richness, and panel B shows the relationship between host functional diversity and symbiont transmission rate. Here, $C = 500$, $H = 50$, $S = 50$, $\sigma = 1$, $c = 0.0001$, $r = 0.1$, $d = 0.08$, $r_s = 1$, and $\phi = 10$.

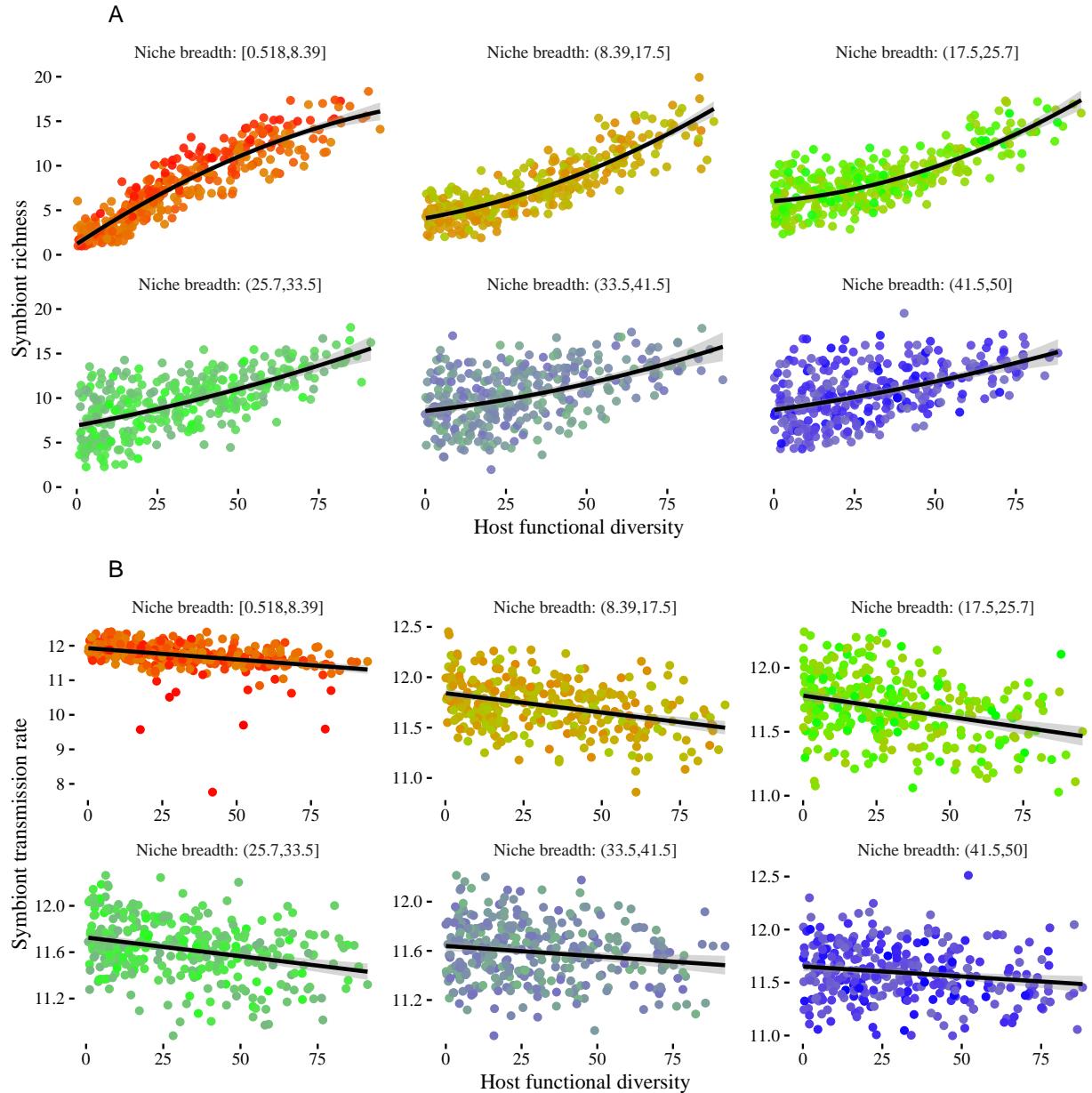


Figure 6.3: Model predictions with varying symbiont niche widths among time series (though constant within). Symbiont niche widths were uniformly generated between $\sigma_{min} = 0.5$ and $\sigma_{max} = 50$, and as before $C = 500$, $H = 50$, $S = 50$, $c = 0.0001$, $r = 0.1$, $d = 0.08$, $r_s = 1$, and $\phi = 10$. Colors represent symbiont niche width and encode the same information as the panels, but aid in comparisons between the symbiont richness plots (panel A) and the symbiont transmission plots (panel B).

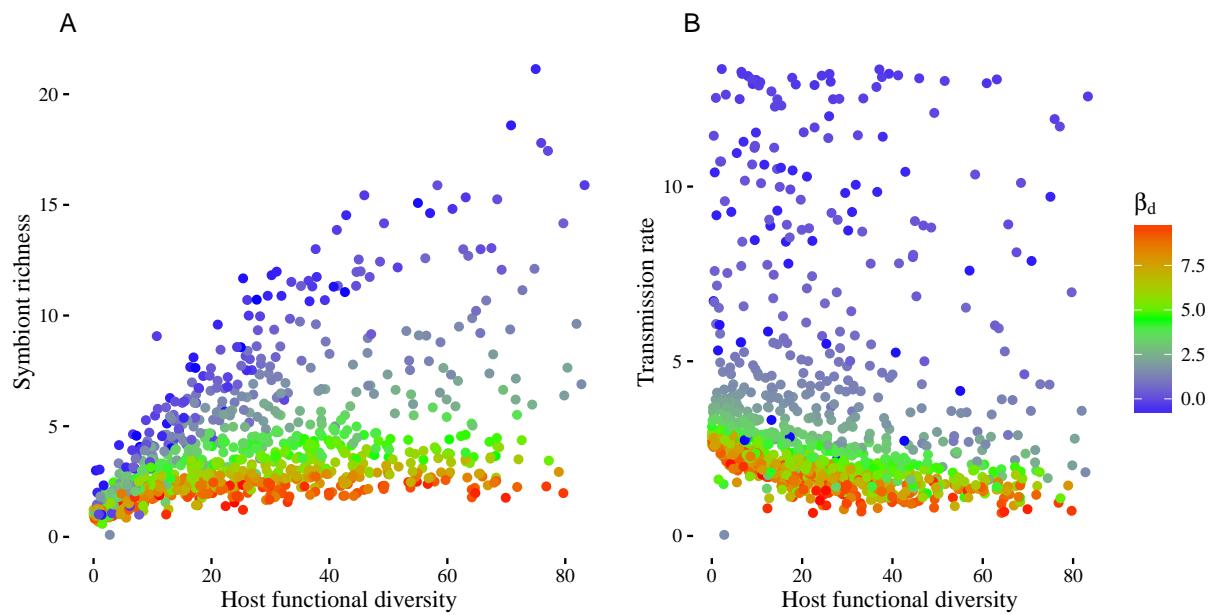


Figure 6.4: Model predictions with varying fitness consequences of infection among time series. Symbiont fitness effects β_d were uniformly generated between -0.99 and 10 . Positive values of β_d imply increased death rates with infection and more parasitic symbionts, while negative values of β_d imply mutualistic symbionts that reduce host death rates with infection. Colors represent the particular values of β_d selected for each simulated time series. Here, $C = 500$, $H = 50$, $S = 50$, $\sigma = 1$, $c = 0.0001$, $r = 0.1$, $d = 0.08$, $r_s = 1$, and $\phi = 10$.

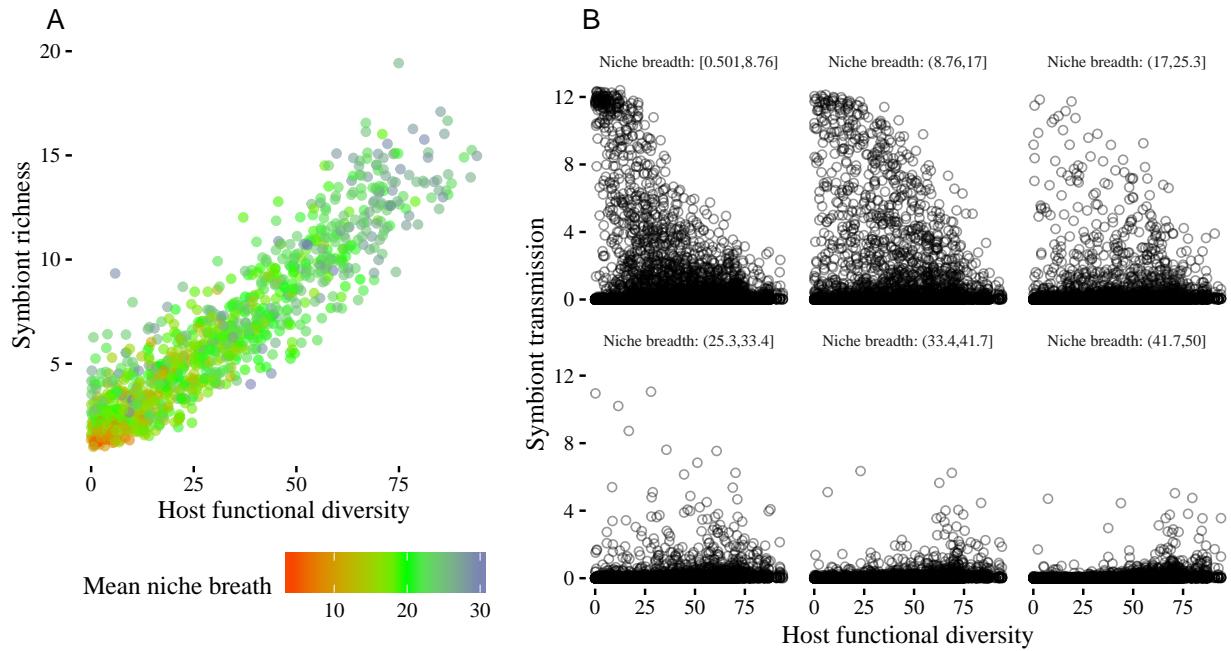


Figure 6.5: Model predictions for mixed niche breadths within simulated time series. Here, symbiont niche breadths vary among symbiont species and are uniformly generated on the interval $(0.5, 50)$. Panel A shows the scaling of host functional diversity with symbiont richness, with the mean symbiont niche breadth in the local community shown in color. Panel B shows the transmission rates for each particular symbiont species by time series combination with panels corresponding to niche width bins. Here, $C = 500$, $H = 50$, $S = 50$, $\sigma = 1$, $c = 0.0001$, $r = 0.1$, $d = 0.08$, $r_s = 1$, and $\phi = 10$.

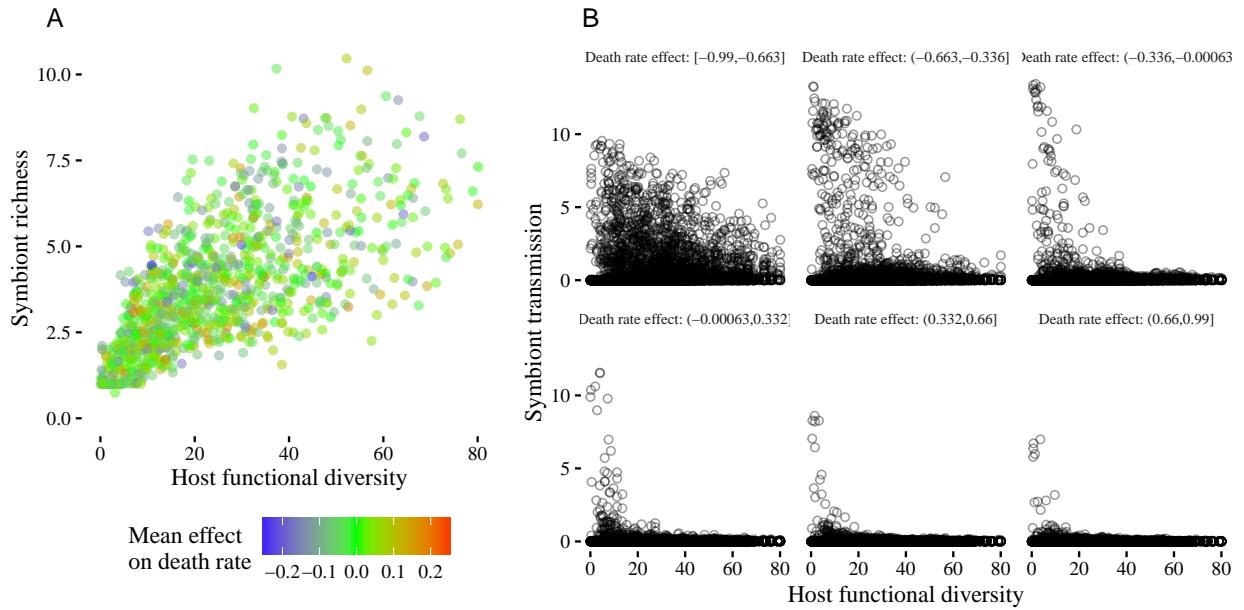


Figure 6.6: Model predictions for mixed fitness consequences within simulated time series. Symbiont effects on the host death rate vary among symbiont species and are uniformly generated on the interval $(-0.99, 0.99)$. Panel A shows the scaling of host functional diversity with symbiont richness, and mean fitness effects are shown in color. Panel B corresponds to the transmission rates for each symbiont species in each time series with panels corresponding to death rate effect bins. If the death rate effect is negative, then infected hosts have lower death rates, surviving longer on average. If positive, hosts die faster. Here, $C = 500$, $H = 50$, $S = 50$, $\sigma = 1$, $c = 0.0001$, $r = 0.1$, $d = 0.08$, $r_s = 1$, and $\phi = 10$.

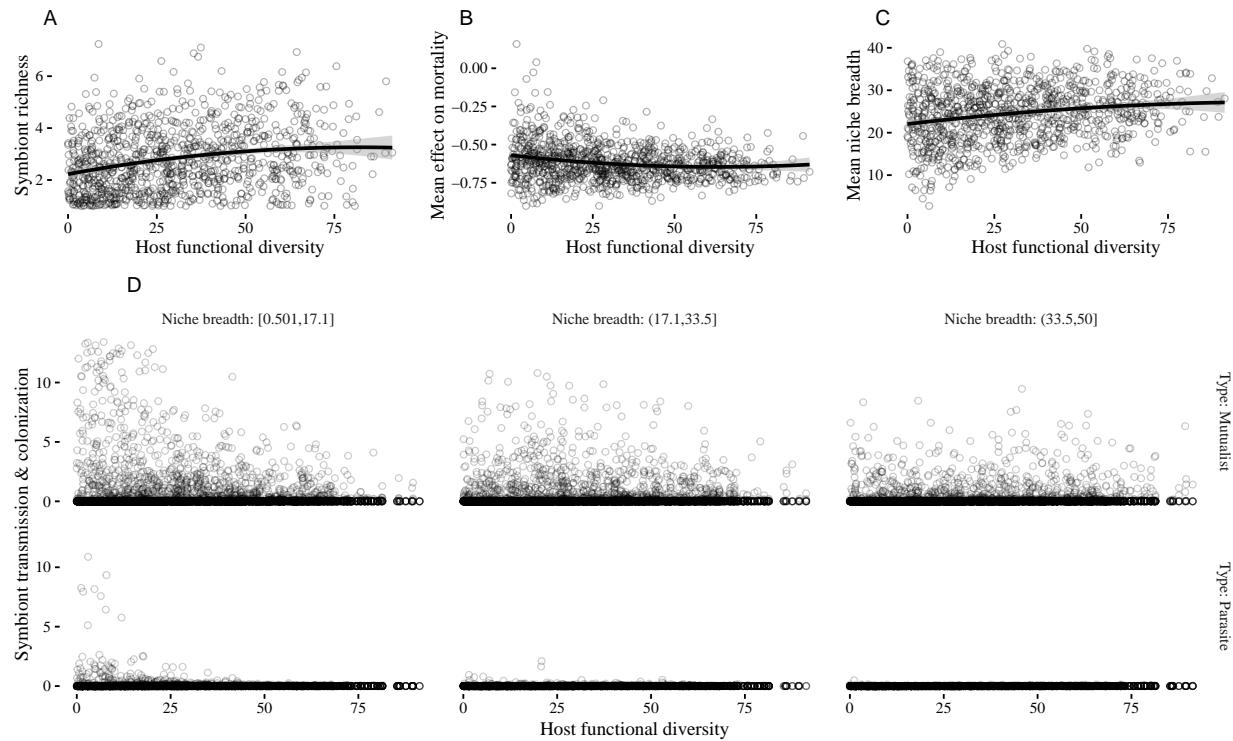


Figure 6.7: Model predictions for mixed symbiont fitness consequences and niche breadths within simulated time series. Symbiont effects on the host death rate vary among symbiont species and are uniformly generated on the interval $(-0.99, 0.99)$, and symbiont niche breadths vary among symbiont species and are uniformly generated on the interval $(0.5, 50)$. Panel A shows the scaling of host functional diversity with symbiont richness. Panel B and C show the relationships between mean mortality effects and niche breadths with host functional diversity. Panel D corresponds to the transmission rates for each symbiont species in each time series with panels corresponding to death rate effect and niche breadth bins. Here, $C = 500$, $H = 50$, $S = 50$, $\sigma = 1$, $c = 0.0001$, $r = 0.1$, $d = 0.08$, $r_s = 1$, and $\phi = 10$.

Bibliography

- (2015). RStan: the R interface to Stan.
- (2015). Stan: A C++ Library for Probability and Sampling, Version 2.8.0.
- Ackerly, D. D. and Cornwell, W. K. (2007). A trait-based approach to community assembly: Partitioning of species trait values into within- and among-community components. *Ecology Letters*, 10(2):135–145.
- Adams, M. J. (2000). Pond permanence and the effects of exotic vertebrates on anurans. *Ecological Applications*, 10(2):559–568.
- Adams, M. J., Pearl, C. A., McCreary, B., Galvan, S. K., Wessell, S. J., Wentz, W. H., Chauncey, W., and Kuehl, A. B. (2009). Short-term effect of cattle exclosures on columbia spotted frog (*Rana luteiventris*) populations and habitat in northeastern Oregon. *Journal of Herpetology*, 43(1):132–138.
- Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B*, 44(2):139–177.
- Aitchison, J. and Ho, C. H. (1989). The multivariate Poisson-log normal distribution. *Biometrika*, 76(4):643–653.
- Alexandre Felizola Diniz-Filho, J. (2001). PHYLOGENETIC AUTOCORRELATION UNDER DISTINCT EVOLUTIONARY PROCESSES. *Evolution*, 55(6):1104–1109.
- Alizon, S. and van Baalen, M. (2008). Multiple infections, immune dynamics, and the evolution of virulence. *The American naturalist*, 172(4):E150–68.
- Allouche, O., Kalyuzhny, M., Moreno-Rueda, G., Pizarro, M., and Kadmon, R. (2012). Area-heterogeneity tradeoff and the diversity of ecological communities., volume 109.
- Almeida, A. M. F., Matsumoto, M. T., Baeza, L. C., Silva, R. B. D., Kleiner, A. A. P., Melhem, M. D., Giannini, M., and Lab Grp, C. (2007). Molecular typing and antifungal susceptibility of clinical sequential isolates of *Cryptococcus neoformans* from São Paulo State, Brazil. *Fems Yeast Research*, 7(1):152–164.
- Alsterberg, C., Eklöf, J. S., Gamfeldt, L., Havenhand, J. N., and Sundbäck, K. (2013). Consumers mediate the effects of experimental ocean acidification and warming on primary producers. *Proceedings of the National Academy of Sciences of the United States of America*, 110(21):8603–8.

- Altizer, S., Dobson, A., Hosseini, P., Hudson, P., Pascual, M., and Rohani, P. (2006). Seasonality and the dynamics of infectious diseases. *Ecology letters*, 9(4):467–484.
- Alvarez, J. A., Dunn, C., and Zuur, A. F. (2003). Response of California Red-legged Frogs to removal of non-native fish. *Transactions of the Western Section of the Wildlife society*, 39:9–12.
- Amarasekare, P., Hoopes, M. F., Mouquet, N., and Holyoak, M. (2004). Mechanisms of coexistence in competitive metacommunities. *The American naturalist*, 164(3):310–326.
- Anderson, R. M., Crombie, J. A., and Grenfell, B. T. (1987). The epidemiology of mumps in the UK: a preliminary study of virus transmission, herd immunity and the potential impact of immunization. *Epidemiology and infection*, 99(01):65–84.
- Anderson, R. M. and May, R. M. (1978). Regulation and stability of host-parasite population interactions: I. regulatory processes. *The Journal of Animal Ecology*, 47(1):219.
- Andersson, G., Berggren, H., Cronberg, G., and Gelin, C. (1978). Effects of planktivorous and benthivorous fish on organisms and water chemistry in eutrophic lakes. *Hydrobiologia*, 59(1):9–15.
- Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4:40–79.
- Ashford, J. R. and Sowden, R. R. (1970). Multi-variate probit analysis. *Biometrics*, 26(3):535–546.
- Ashford, R. W. (2003). When is a reservoir not a reservoir? *Emerging infectious diseases*, 9(11):1495–1496.
- Bailey, L. L., MacKenzie, D. I., and Nichols, J. D. (2014). Advances and applications of occupancy models. *Methods in Ecology and Evolution*, 5(12):1269–1279.
- Barbara, S., Lucia, S., Biology, E., Gerber, L. R., Botsford, L. W., Hastings, A., Possingham, H. P., Gaines, S. D., Palumbi, S. R., and Andelman, S. (2003). Population Models for Marine Reserve Design: a Retrospective and Prospective Synthesis. *Ecological Applications*, 13(sp1):47–64.
- Bartelt, G. A., Pardee, J., and Thiede, K. (2003). *Environmental impact statement on rules to eradicate chronic* w. Wisconsin Department of Natural Resources, Bureau of Integrated Science Services.
- Bashan, Y. (1998). Inoculants of plant growth-promoting bacteria for use in agriculture. *Biotechnology Advances*, 16(4):729–770.
- Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin*, pages 47–53.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons, Inc.
- Bolzoni, L., De Leo, G. a., Gatto, M., and Dobson, A. P. (2008). Body-size scaling in an SEI model of wildlife diseases. *Theoretical population biology*, 73(3):374–82.
- Bordes, F., Morand, S., Kelt, D. a., and Van Vuren, D. H. (2009). Home range and parasite diversity in mammals. *The American naturalist*, 173(4):467–74.

- Bowers, R. G. and Turner, J. (1997). Community structure and the interplay between interspecific infection and competition. *Journal of Theoretical Biology*, 187(1):95–109.
- Boyle, S. a. and Smith, A. T. (2010). Can landscape and species characteristics predict primate presence in forest fragments in the Brazilian Amazon? *Biological Conservation*, 143(5):1134–1143.
- Brodman, R., Ogger, J., Bogard, T., Long, A. J., Pulver, R. A., Mancuso, K., and Falk, D. (2003). Multivariate analyses of the influences of water chemistry and habitat parameters on the abundances of pond-breeding amphibians. *Journal of Freshwater Ecology*, 18(3):425–436.
- Broms, K. M., Hooten, M. B., Johnson, D. S., Altwegg, R., and Conquest, L. L. (2015). Dynamic occupancy models for explicit colonization processes. *Ecology*, in press:150716143727003.
- Brooks, S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455.
- Bruno, J. F. and Cardinale, B. J. (2008). Cascading effects of predator richness. *Frontiers in Ecology and the Environment*, 6(10):539–546.
- Buckley, L. B. and Jetz, W. (2007). Insularity and the determinants of lizard population density. *Ecology letters*, 10(6):481–9.
- Burne, M. R. and Griffin, C. R. (2005). Habitat associations of pool-breeding amphibians in eastern Massachusetts, USA. *Wetlands Ecology and Management*, 13(3):247–259.
- Cadotte, M. W., Carscadden, K., and Mirochnick, N. (2011). Beyond species: functional diversity and the maintenance of ecological processes and services. *Journal of Applied Ecology*, 48(5):1079–1087.
- Cardillo, M. and Bromham, L. (2001). Body size and risk of extinction in Australian mammals. *Conservation Biology*, 15(5):1435–1440.
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., Mace, G. M., Tilman, D., Wardle, D. a., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., and Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486(7401):59–67.
- Chen, L. and Zhou, S. (2015). A Combination of Species Evenness and Functional Diversity Is the Best Predictor of Disease Risk in Multihost Communities. *The American Naturalist*, 186(6):000–000.
- Choo, K., Williams, P. D., and Day, T. (2003). Host mortality, predation and the evolution of parasite virulence. *Ecology letters*, 6(4):310–315.
- Civitello, D. J., Cohen, J., Fatima, H., Halstead, N. T., Liriano, J., McMahon, T. a., Ortega, C. N., Sauer, E. L., Sehgal, T., Young, S., and Rohr, J. R. (2015). Biodiversity inhibits parasites: Broad evidence for the dilution effect. *Proceedings of the National Academy of Sciences*, 112(28):201506279.
- Clark, J. S., Gelfand, A. E., Woodall, C. W., and Zhu, K. (2014). More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications*, 24(5):990–999.

- Clough, Y. (2012). A generalized approach to modeling and estimating indirect effects in ecology. *Ecology*, 93(8):1809–1815.
- Collins, J. (1979). Intrapopulation variation in the body size at metamorphosis and timing of metamorphosis in the bullfrog, *< i>Rana catesbeiana</i>*. *Ecology*, 60(4):738–749.
- Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., and Piatek, R. (2014). Bayesian Exploratory Factor Analysis. *Journal of econometrics*, 183(1):31–57.
- Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J., and Relman, D. A. (2012). The application of ecological theory towards an understanding of the human microbiome. *Science*, 336(6086):1255–1262.
- Creech, T. G., Cross, P. C., Scurlock, B. M., Maichak, E. J., Rogerson, J. D., Henningsen, J. C., and Creel, S. (2012). Effects of low-density feeding on elk-fetus contact rates on Wyoming feedgrounds. *The Journal of Wildlife Management*, 76(5):877–886.
- Cressie, N., Calder, C. A., Clark, J. S., Ver Hoef, J. M., and Wikle, C. K. (2009). Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, 19(3):553–570.
- Cronin, J. P., Welsh, M. E., Dekkers, M. G., Abercrombie, S. T., and Mitchell, C. E. (2010). Host physiological phenotype explains pathogen reservoir potential. *Ecology letters*, 13(10):1221–32.
- Cross, P. C., Edwards, W. H., Scurlock, B. M., Maichak, E. J., and Rogerson, J. D. (2007). Effects of management and climate on elk brucellosis in the Greater Yellowstone Ecosystem. *Ecological applications : a publication of the Ecological Society of America*, 17(4):957–64.
- Crump, M. and Bury, R. (1994). Visual encounter surveys. In Heyer, W., Donnelly, M., McDiarmid, R., Hayek, L., and Foster, M., editors, *Measuring and monitoring biological diversity, standard methods for amphibians*, pages 84–92. Smithsonian Institution Press, Washington DC.
- Dail, D. and Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, 67:577–587.
- Damuth, J. (1981). Population density and body size in mammals. *Nature*, 290:699–700.
- Davidson, A. D., Hamilton, M. J., Boyer, A. G., Brown, J. H., and Ceballos, G. (2009). Multiple ecological pathways to extinction in mammals. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26):10702–5.
- de Castro, F. and Bolker, B. (2004). Mechanisms of disease-induced extinction. *Ecology Letters*, 8(1):117–126.
- Deem, S. L., Karesh, W. B., and Weisman, W. (2001). Putting Theory into Practice: Wildlife Health in Conservation. *Conservation Biology*, 15(5):1224–1233.
- Dobson, A. (2004). Population dynamics of pathogens with multiple host species. *The American naturalist*, 164 Suppl(November 2004):S64—78.
- Dobson, A., Lodge, D., Alder, J., and Cumming, G. (2006). Habitat loss, trophic collapse, and the decline of ecosystem services. *Ecology*, 87(8):1915–1924.

- Dorazio, R. M. and Connor, E. F. (2014). Estimating abundances of interacting species using morphological traits, foraging guilds, and habitat. *PLoS ONE*, 9(4):e94323.
- Dorazio, R. M., Kéry, M., Royle, J. A., and Plattner, M. (2010). Models for inference in dynamic metacommunity systems. *Ecology*, 91(8):2466–2475.
- Driscoll, D. A. and Lindenmayer, D. B. (2012). Framework to improve the application of theory in ecology and conservation. *Ecological Monographs*, 82(1):129–147.
- Duffy, J. E. (2003). Biodiversity loss, trophic skew and ecosystem functioning. *Ecology Letters*, 6(8):680–687.
- Dunn, R. R., Davies, T. J., Harris, N. C., and Gavin, M. C. (2010). Global drivers of human pathogen richness and prevalence. *Proceedings. Biological sciences / The Royal Society*, 277(1694):2587–2595.
- Earl, J. E. and Whiteman, H. H. (2009). Effects of pulsed nitrate exposure on amphibian development. *Environmental toxicology and chemistry*, 28(6):1331–1337.
- Ebert, D., Lipsitch, M., and Mangin, K. L. (2000). The effect of parasites on host population density and extinction: experimental epidemiology with *Daphnia* and six microparasites. *The American Naturalist*, 156(5):459–477.
- Ebinger, M., Cross, P., Wallen, R., White, P. J., and Treanor, J. (2011). Simulating sterilization, vaccination, and test-and-remove as brucellosis control measures in bison. *Ecological Applications*, 21(8):2944–2959.
- Egan, R. and Paton, P. (2004). Within-pond parameters affecting oviposition by wood frogs and spotted salamanders. *Wetlands*, 24(1):1–13.
- Elias, D., Akuffo, H., and Britton, S. (2006). Helminthes could influence the outcome of vaccines against TB in the tropics. *Parasite Immunology*, 28(10):507–513.
- Ezenwa, V. O., Etienne, R. S., Luikart, G., Beja-Pereira, A., and Jolles, A. E. (2010). Hidden consequences of living in a wormy world: nematode-induced immune suppression facilitates tuberculosis invasion in African buffalo. *The American naturalist*, 176(5):613–24.
- Felizola Diniz Filho, J. A., Rangel, T. F., Santos, T., and Mauricio Bini, L. (2012). Exploring Patterns of Interspecific Variation in Quantitative Traits Using Sequential Phylogenetic Eigenvector Regressions. *Evolution*, 66(4):1079–1090.
- Felsenstein, J. (1985). Phylogenies and the comparative method. *The American Naturalist*, 125(1):1–15.
- Fierer, J., Burton, D. W., Haghghi, P., and Deftos, L. J. (2012). Hypercalcemia in Disseminated Coccidioidomycosis: Expression of Parathyroid Hormone-Related Peptide Is Characteristic of Granulomatous Inflammation. *Clinical Infectious Diseases*, 55(7):E61–E66.
- Fisher, R. N. and Shaffer, H. B. (1996). The Decline of amphibians in California's Great Central Valley. *Conservation Biology*, 10(5):1387–1397.
- Fleischner, T. L. (1994). Ecological costs of livestock grazing in western North America.

- Frank, K. T. and Leggett, W. C. (1994). Fisheries ecology in the context of ecological and evolutionary theory. *Annual Review of Ecology and Systematics*, pages 401–422.
- Freda, J. and Dunson, W. (1986). Effects of low pH and other chemical variables on the local distribution of amphibians. *Copeia*, 1986(2):454–466.
- Gandon, S. and Day, T. (2007). The evolutionary epidemiology of vaccination. *Journal of the Royal Society Interface*, 4(16):803–817.
- Garone, P. (2011). *The Fall and Rise of the Wetlands of California's Great Central Valley*. University of California Press, 1 edition.
- Gaston, K. J. and Fuller, R. a. (2008). Commonness, population depletion and conservation biology. *Trends in ecology & evolution*, 23(1):14–9.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.
- Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016.
- Gelman, A. and Pardoe, I. (2006). Bayesian Measures of Explained Variance and Pooling in Multilevel (Hierarchical) Models. *Technometrics*, 48(2):241–251.
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–511.
- Gibbons, P., Zammit, C., Youngentob, K., Possingham, H. P., Lindenmayer, D. B., Bekessy, S., Burgman, M., Colyvan, M., Considine, M., Felton, A., Hobbs, R. J., Hurley, K., McAlpine, C., McCarthy, M. a., Moore, J., Robinson, D., Salt, D., and Wintle, B. (2008). Some practical suggestions for improving engagement between researchers and policy-makers in natural resource management. *Ecological Management & Restoration*, 9(3):182–186.
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434.
- Gimenez, O., Anker-Nilssen, T., and Grosbois, V. (2012). Exploring causal pathways in demographic parameter variation: path analysis of mark-recapture data. *Methods in Ecology and Evolution*, 3(2):427–432.
- Gog, J. R., Pellis, L., Wood, J. L. N., McLean, A. R., Arinaminpathy, N., and Lloyd-Smith, J. O. (2014). Seven challenges in modeling pathogen dynamics within-host and across scales. *Epidemics*, 10:45–48.
- Gómez, J. M., Nunn, C. L., and Verdú, M. (2013). Centrality in primate-parasite networks reveals the potential for the transmission of emerging infectious diseases to humans. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19):7738–41.
- Gottdenker, N. L., Chaves, L. F., Calzada, J. E., Saldaña, A., and Carroll, C. R. (2012). Host life history strategy, species diversity, and habitat influence Trypanosoma cruzi vector infection in Changing landscapes. *PLoS neglected tropical diseases*, 6(11):e1884.

- Grace, J. B. (2006). *Structural Equation Modeling and Natural Systems*. Cambridge Univ Press.
- Grace, J. B., Adler, P. B., Stanley Harpole, W., Borer, E. T., and Seabloom, E. W. (2014). Causal networks clarify productivity-richness interrelations, bivariate plots do not. *Functional Ecology*, 28(4):787–798.
- Grace, J. B., Schoolmaster, D. R., Guntenspergen, G. R., Little, A. M., Mitchell, B. R., Miller, K. M., and Schweiger, E. W. (2012). Guidelines for a graph-theoretic implementation of structural equation modeling. *Ecosphere*, 3(8):1–44.
- Gregory, R., Ohlson, D., and Arvai, J. (2006). Deconstructing adaptive management: criteria for applications to environmental management. *Ecological applications : a publication of the Ecological Society of America*, 16(6):2411–25.
- Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L. N., Daly, J. M., Mumford, J. A., and Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, 303(5656):327–32.
- Gross, J. E. and Miller, M. W. (2001). Chronic wasting disease in mule deer: disease dynamics and control. *The Journal of wildlife management*, pages 205–215.
- Gross, N., Kunstler, G., Liancourt, P., De Bello, F., Suding, K. N., and Lavorel, S. (2009). Linking individual response to biotic interactions with community structure: a trait-based framework. *Functional Ecology*, 23(6):1167–1178.
- Guénard, G., Legendre, P., and Peres-Neto, P. (2013). Phylogenetic eigenvector maps: A framework to model and predict species traits. *Methods in Ecology and Evolution*, 4(12):1120–1131.
- Guisan, A. and Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8:993–1009.
- Hansen, T. F. and Martins, E. P. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*, 50(4):1404–1417.
- Hanski, I. (1991). Single-species metapopulation dynamics: concepts, models and observations. *Biological Journal of the Linnean Society*, 42:17–38.
- Harris, N. C. and Dunn, R. R. (2010). Using host associations to predict spatial patterns in the species richness of the parasites of North American carnivores. *Ecology Letters*, 13(11):1411–1418.
- Harris, R., Lauer, a., Simon, M., Banning, J., and Alford, R. (2009). Addition of antifungal skin bacteria to salamanders ameliorates the effects of chytridiomycosis. *Diseases of Aquatic Organisms*, 83:11–16.
- Hawley, D. M. and Altizer, S. M. (2010). Disease ecology meets ecological immunology: understanding the links between organismal immunity and infection dynamics in natural populations. *Functional Ecology*, pages no–no.

- Haydon, D. T., Randall, D. a., Matthews, L., Knobel, D. L., Tallents, L. a., Gravenor, M. B., Williams, S. D., Pollinger, J. P., Cleaveland, S., Woolhouse, M. E. J., Sillero-Zubiri, C., Marino, J., Macdonald, D. W., and Laurenson, M. K. (2006). Low-coverage vaccination strategies for the conservation of endangered species. *Nature*, 443(7112):692–5.
- He, F. and Gaston, K. J. (2003). Occupancy, spatial variance, and the abundance of species. *The American naturalist*, 162(3):366–375.
- He, F., Gaston, K. J., and Wu, J. (2002). On species occupancy-abundance models. *Ecoscience*, 9(1):119–126.
- Hebblewhite, M., Merrill, E. H., and McDonald, T. L. (2005). Spatial decomposition of predation risk using resource selection functions: an example in a wolf-elk predator-prey system. *Oikos*, 111(1):101–111.
- Heisey, D. M., Osnas, E. E., Cross, P. C., Joly, D. O., Langenberg, J. a., and Miller, M. W. (2010). Linking process to pattern: estimating spatiotemporal dynamics of a wildlife epidemic from cross-sectional data. *Ecological Monographs*, 80(2):221–240.
- Henle, K., Davies, K. F., Kleyer, M., Margules, C., and Settele, J. (2004). Predictors of Species Sensitivity to Fragmentation. *Biodiversity and Conservation*, 13(1):207–251.
- Henrikson, L., Nyman, H. G., Oscarson, H. G., and Stenson, J. A. E. (1980). Trophic changes, without changes in the external nutrient loading. *Hydrobiologia*, 68(3):257–263.
- Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn Sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research*, 15(1):1593–1623.
- Hoffmann, A. A., Montgomery, B. L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P. H., Muzzi, F., Greenfield, M., Durkan, M., Leong, Y. S., Dong, Y., and Others (2011). Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature*, 476(7361):454–457.
- Hogan, J. W. and Tchernis, R. (2004). Bayesian Factor Analysis for Spatially Correlated Data, With Application to Summarizing Area-Level Material Deprivation From Census Data. *Journal of the American Statistical Association*, 99(466):314–324.
- Holling, C. S. and Others (1978). Adaptive environmental assessment and management. *Adaptive environmental assessment and management*.
- Holt, R. D. and Keitt, T. H. (2000). Alternative causes for range limits: a metapopulation perspective. *Ecology Letters*, 3(1):41–47.
- Holt, R. D. and Roy, M. (2007). Predation can increase the prevalence of infectious disease. *The American naturalist*, 169(5):690–699.
- Hoverman, J. T., Gray, M. J., and Miller, D. L. (2010). Anuran susceptibilities to ranaviruses: role of species identity, exposure route, and a novel virus isolate. *Diseases of Aquatic Organisms*, 89(2):97–107.
- Hoverman, J. T., Hoye, B. J., and Johnson, P. T. J. (2013). Does timing matter? How priority effects influence the outcome of parasite interactions within hosts. *Oecologia*, 173(4):1471–1480.

- Huang, Y., Mi, Z. Q., Zhuang, L., Ma, M. J., An, X. P., Liu, W., Cao, W. C., and Tong, Y. G. (2013a). Presence of entomobirnaviruses in Chinese mosquitoes in the absence of Dengue virus co-infection. *Journal of General Virology*, 94:663–667.
- Huang, Z. Y. X., de Boer, W. F., van Langevelde, F., Olson, V., Blackburn, T. M., and Prins, H. H. T. (2013b). Species' life-history traits explain interspecific variation in reservoir competence: a possible mechanism underlying the dilution effect. *PloS one*, 8(1):e54341.
- Hudson, P. J., Rizzoli, A., Grenfell, B. T., Heesterbeek, H., and Dobson, A. P. (2002). *The ecology of wildlife diseases*. Oxford University Press Oxford.
- Hyatt, A. D., Boyle, D. G., Olsen, V., Boyle, D. B., Berger, L., Obendorf, D., Dalton, A., Kriger, K., Hero, M., Hines, H., Phillott, R., Campbell, R., Marantelli, G., Gleason, F., and Colling, A. (2007). Diagnostic assays and sampling protocols for the detection of Batrachochytrium dendrobatidis. *Diseases of Aquatic Organisms*, 73(3):175–192.
- Irianto, A. and Austin, B. (2002). Probiotics in aquaculture. *Journal of fish diseases*, 25(11):633–642.
- Jansen, A. and Healey, M. (2003). Frog communities and wetland condition: relationships with grazing by domestic livestock along an Australian floodplain river. *Biological Conservation*, 109(2):207–219.
- Johnson, A. R. (2012). Avoiding Environmental Catastrophes: Varieties of Principled Precaution. *Ecology and Society*, 17(3).
- Johnson, P. T. J. and Buller, I. D. (2011). Parasite competition hidden by correlated coinfection: using surveys and experiments to understand parasite interactions. *Ecology*, 92(3):535–541.
- Johnson, P. T. J., de Roode, J. C., and Fenton, A. (2015). Why infectious disease research needs community ecology. *Science*, 349(6252):1259504.
- Johnson, P. T. J. and Hoverman, J. T. (2012). Parasite diversity and coinfection determine pathogen infection success and host fitness. *Proceedings of the National Academy of Sciences of the United States of America*, 109(23):9006–11.
- Johnson, P. T. J., Preston, D. L., Hoverman, J. T., and Richgels, K. L. D. (2013). Biodiversity decreases disease through predictable changes in host community competence. *Nature*, 494(7436):230–233.
- Jones, K. E., Patel, N. G., Levy, M. a., Storeygard, A., Balk, D., Gittleman, J. L., and Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, 451(7181):990–3.
- Joseph, M. B., Mihaljevic, J. R., Orlofske, S. a., and Paull, S. H. (2013). Does life history mediate changing disease risk when communities disassemble? *Ecology Letters*, 16(11):1405–1412.
- Joseph, M. B., Preston, D. L., and Johnson, P. T. J. (2015). Integrating occupancy models and structural equation models to understand species occurrence. *Ecology*, (in press).
- Kaltz, O. and Shykoff, J. a. (1998). Local adaptation in hostparasite systems. *Heredity*, 81(4):361–370.

- Kats, L. B., Petranka, J. W., and Sih, A. (1988). Antipredator defenses and the persistence of amphibian larvae with fishes.
- Kauffman, J., Krueger, W., and Vavra, M. (1983). Effects of late season cattle grazing on riparian plant communities. *Journal of Range Management*, 36(6):685–691.
- Keeling, M. J., Woolhouse, M. E., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J., and Grenfell, B. T. (2001). Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science (New York, N.Y.)*, 294(5543):813–817.
- Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., Hudson, P., Jolles, A., Jones, K. E., Mitchell, C. E., Myers, S. S., Bogich, T., and Ostfeld, R. S. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature*, 468(7324):647–652.
- Keesing, F., Holt, R. D., and Ostfeld, R. S. (2006). Effects of species diversity on disease risk. *Ecology letters*, 9(4):485–98.
- Knutson, M. G. M. G., Richardson, W. B. W. B., Reineke, D. M., Gray, B. R., Parmelee, J. R., and Weick, S. E. (2004). Agricultural ponds support amphibian populations. *Ecological Applications*, 14(3):669–684.
- Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Adaptive Computation and Machine Learning. MIT Press.
- Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R., and Bunning, M. (2003). Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerging infectious diseases*, 9(3):311–22.
- Kruse, K. and Francis, M. (1977). A predation deterrent in larvae of the bullfrog, *< i>Rana catesbeiana</i>*. *Transactions of the American Fisheries Society*, 106:248–252.
- Kruse, K. P. and Stone, B. M. (1984). Largemouth bass (*< i>Micropterus salmoides</i>*) learn to avoid feeding on toad (*< i>Bufo</i>*) tadpoles. *Animal Behaviour*, 32:1035–1039.
- Kucukelbir, A., Ranganath, R., Gelman, A., and Blei, D. M. (2015). Automatic Variational Inference in Stan. pages 1–22.
- Kuris, A. M., Blaustein, A. R., and Alio, J. J. (1980). Hosts as Islands. *American naturalist*, 116(4):570–586.
- Lachish, S., Gopalaswamy, A. M., Knowles, S. C. L., and Sheldon, B. C. (2012). Site-occupancy modelling as a novel framework for assessing test sensitivity and estimating wildlife disease prevalence from imperfect diagnostic tests. *Methods in Ecology and Evolution*, 3(2):339–348.
- Lafferty, K. D. (2012). Biodiversity loss decreases parasite diversity: theory and patterns. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367:2814–2827.
- Lafferty, K. D. and Gerber, L. R. (2002). Good Medicine for Conservation Biology: the Intersection of Epidemiology and Conservation Theory. *Conservation Biology*, 16(3):593–604.
- Lajeunesse, M. J. and Forbes, M. R. (2002). Host range and local parasite adaptation. *Proceedings of the Royal Society B Biological Sciences*, 269(1492):703–710.

- Lam, B. a., Walke, J. B., Vredenburg, V. T., and Harris, R. N. (2010). Proportion of individuals with anti-Batrachochytrium dendrobatidis skin bacteria is associated with population persistence in the frog *Rana muscosa*. *Biological Conservation*, 143(2):529–531.
- Lamb, E. G., Mengersen, K. L., and Stewart, K. J. (2014). Spatially Explicit Structural Equation Modeling. *Ecology*, 95(9):2434–2442.
- Lande, R. (1993). Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes Author (s): Russell Lande Source : The American Naturalist , Vol . 142 , No . 6 (Dec ., 1993), pp . 911-927 Published by : The University of Chicag. *The American Naturalist*, 142(6):911–927.
- Langwig, K. E., Frick, W. F., Bried, J. T., Hicks, A. C., Kunz, T. H., and Kilpatrick, a. M. (2012). Sociality, density-dependence and microclimates determine the persistence of populations suffering from a novel fungal disease, white-nose syndrome. *Ecology letters*, 15(9):1050–7.
- Lannoo, M. J. (2005). *Amphibian Declines: The Conservation Status of United States Species*. University of California Press.
- Lawler, S. P., Dritz, D., Strange, T., and Holyoak, M. (1999). Effects of introduced mosquitofish and bullfrogs on the threatened California red-legged frog. *Conservation Biology*, 13:613–622.
- Lee, K. a. (2006). Linking immune defenses and life history at the levels of the individual and the species. *Integrative and Comparative Biology*, 46(6):1000–1015.
- Lee, S. (2007). *Structural Equation Modeling: A Bayesian Approach*. Wiley.
- Lee, T. M. and Jetz, W. (2011). Unravelling the structure of species extinction risk for predictive conservation science Unravelling the structure of species extinction risk for predictive conservation science. (October 2010):1329–1338.
- Leo, G. D. and Dobson, A. (1996). Allometry and simple epidemic models for microparasites. *Nature*, 379:720–722.
- Levi, T., Kilpatrick, A. M., Mangel, M., and Wilmers, C. C. (2012). Deer , predators , and the emergence of Lyme disease. 2012.
- Levins, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15(3):237–240.
- Little, R. J. A. and Rubin, D. B. (2014). *Statistical analysis with missing data*. John Wiley & Sons.
- Lively, C. M., de Roode, J. C., Duffy, M. a., Graham, A. L., and Koskella, B. (2014). Interesting Open Questions in Disease Ecology and Evolution*. *The American Naturalist*, 184(August 2014):S000–S000.
- Lively, C. M. and Dybdahl, M. F. (2000). Parasite adaptation to locally common host genotypes. *Nature*, 405(6787):679–81.
- Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359.

- Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica*, 14:41–67.
- López, G., López-Parra, M., Fernández, L., Martínez-Granados, C., Martínez, F., Meli, M. L., Gil-Sánchez, J. M., Viqueira, N., Díaz-Portero, M. a., Cadenas, R., Lutz, H., Vargas, a., and Simón, M. a. (2009). Management measures to control a feline leukemia virus outbreak in the endangered Iberian lynx. *Animal Conservation*, 12(3):173–182.
- MacArthur, R., Diamond, J., and Karr, J. (1972). Density compensation in island faunas. *Ecology*, 53(2):330–342.
- MacKenzie, D., Nichols, J., and Lachman, G. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.
- MacKenzie, D. I. and Bailey, L. L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(3):300–318.
- MacKenzie, D. I., Bailey, L. L., Hines, J. E., and Nichols, J. D. (2011). An integrated model of habitat and species occurrence dynamics. *Methods in Ecology and Evolution*, 2(6):612–622.
- Mackinnon, M. J., Gandon, S., and Read, A. F. (2008). Virulence evolution in response to vaccination: the case of malaria. *Vaccine*, 26:C42—C52.
- Maichak, E. J., Scurlock, B. M., Rogerson, J. D., Meadows, L. L., Barbknecht, A. E., Edwards, W. H., and Cross, P. C. (2009). Effects of management, behavior, and scavenging on risk of brucellosis transmission in elk of western Wyoming. *Journal of Wildlife Diseases*, 45(2):398–410.
- Martin, L. B., Hasselquist, D., and Wikelski, M. (2006). Investment in immune defense is linked to pace of life in house sparrows. *Oecologia*, 147(4):565–75.
- Marty, J. T. (2005). Effects of cattle grazing on diversity in ephemeral wetlands. *Conservation Biology*, 19:1626–1632.
- May, R. M. and Anderson, R. M. (1979). Population biology of infectious diseases: Part II. *Nature*, 280(5722):455–461.
- McCallum, H. (2012). Disease and the dynamics of extinction. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1604):2828–2839.
- McCallum, H., Barlow, N., and Hone, J. (2001). How should pathogen transmission be modelled? *Trends in ecology & evolution*, 16(6):295–300.
- McDonald-Madden, E., Probert, W. J. M., Hauser, C. E., Runge, M. C., Possingham, H. P., Jones, M. E., Moore, J. L., Rout, T. M., Veske, P. a., and Wintle, B. a. (2010). Active adaptive conservation of threatened species in the face of uncertainty. *Ecological applications : a publication of the Ecological Society of America*, 20(5):1476–89.
- McFarland, L. V. (2006). Meta-analysis of probiotics for the prevention of antibiotic associated diarrhea and the treatment of Clostridium difficile disease. *The American journal of gastroenterology*, 101(4):812–822.
- Meehan, T. D., Jetz, W., and Brown, J. H. (2004). Energetic determinants of abundance in winter landbird communities. *Ecology Letters*, 7(7):532–537.

- Michalski, F. and Peres, C. a. (2005). Anthropogenic determinants of primate and carnivore local extinctions in a fragmented forest landscape of southern Amazonia. *Biological Conservation*, 124(3):383–396.
- Mideo, N., Alizon, S., and Day, T. (2008). Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases. *Trends in Ecology and Evolution*, 23(9):511–517.
- Mihaljevic, J. R. (2012). Linking metacommunity theory and symbiont evolutionary ecology. *Trends in Ecology and Evolution*, 27(6):323–329.
- Mihaljevic, J. R., Joseph, M. B., Orlofske, S. A., and Paull, S. H. (2014). The scaling of host density with richness affects the direction, shape, and detectability of diversity-disease relationships. *PloS one*, 9(5):e97812.
- Miller, D. A. W., Talley, B. L., Lips, K. R., and Campbell Grant, E. H. (2012). Estimating patterns and drivers of infection prevalence and intensity when detection is imperfect and sampling error occurs. *Methods in Ecology and Evolution*, 3(5):850–859.
- Miller, Z. J. (2012). Fungal pathogen species richness: why do some plant species have more pathogens than others? *The American Naturalist*, 179(2):282–92.
- Mitchell, C. E., Agrawal, A. A., Bever, J. D., Gilbert, G. S., Hufbauer, R. A., Klironomos, J. N., Maron, J. L., Morris, W. F., Parker, I. M., Power, A. G., Seabloom, E. W., Torchin, M. E., and Vazquez, D. P. (2006). Biotic interactions and plant invasions. *Ecology Letters*, 9(6):726–740.
- Montarry, J., Glais, I., Corbiere, R., and Andrivon, D. (2008). Adaptation to the most abundant host genotype in an agricultural plant-pathogen system—potato late blight. *Journal of Evolutionary Biology*, 21(5):1397–1407.
- Morais, A. R., Siqueira, M. N., Lemes, P., Maciel, N. M., De Marco, P., and Brito, D. (2013). Unraveling the conservation status of data deficient species. *Biological Conservation*, 166:98–102.
- Moran, N. A. and Baumann, P. (2000). Bacterial endosymbionts in animals. *Current Opinion in Microbiology*, 3(3):270–275.
- Myhrvold, N. P., Baldridge, E., Chan, B., Sivam, D., Freeman, D. L., and Ernest, S. K. (2015). An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles. *Ecology*, 96(11):3109.
- Nakagawa, S. and Freckleton, R. P. (2008). Missing inaction: the dangers of ignoring missing data. *Trends in Ecology and Evolution*, 23(11):592–596.
- Neal, R. M. (1998). Regression and Classification Using Gaussian Process Priors. *Bayesian Statistics*, 6:475–501.
- Nee, S., Harvey, P. H., and May, R. M. (1991). Lifting the Veil on Abundance Patterns. *Proceedings of the Royal Society B: Biological Sciences*, 243(1307):161–163.
- Nunn, C. L., Altizer, S. M., Sechrest, W., and Cunningham, A. A. (2005). Latitudinal gradients of parasite species richness in primates. *Diversity and Distributions*, 11(3):249–256.

- Olden, J. D., Hogan, Z. S., and Zandeen, M. J. V. (2007). Small fish, big fish, red fish, blue fish: size-biased extinction risk of the world's freshwater and marine fishes. *Global Ecology and Biogeography*, 16(6):694–701.
- ORegan, S. M., Vinson, J. E., and Park, A. W. (2015). Interspecific Contact and Competition May Affect the Strength and Direction of Disease-Diversity Relationships for Directly Transmitted Microparasites. *The American Naturalist*, 186(4):480–494.
- Ospina, R. and Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6):1609–1623.
- Ostfeld, R. S. and Keesing, F. (2000). Biodiversity series: the function of biodiversity in the ecology of vector-borne zoonotic diseases. *Canadian Journal of Zoology*, 78(12):2061–2078.
- Ostfeld, R. S. and Keesing, F. (2012). Effects of Host Diversity on Infectious Disease. *Annual Review of Ecology, Evolution, and Systematics*, 43(1):157–182.
- Ostfeld, R. S., Keesing, F., and Eviner, V. T. (2010). *Infectious disease ecology: effects of ecosystems on disease and of disease on ecosystems*. Princeton University Press.
- Ostfeld, R. S. and LoGiudice, K. (2003). Community Disassembly, Biodiversity Loss, and the Erosion of an Ecosystem Service. *Ecology*, 84(6):1421–1427.
- Ostfeld, R. S. R. S. and Holt, R. D. R. D. (2004). Are predators good for your health? Evaluating evidence for top-down regulation of zoonotic disease reservoirs. *Frontiers in Ecology and the Environment*, 2(1):13–20.
- Ovaskainen, O., Hottola, J., and Shtonen, J. (2010). Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, 91(9):2514–2521.
- Pacala, S. and Roughgarden, J. (1982). Spatial heterogeneity and interspecific competition. *Theoretical Population Biology*, 21:92–113.
- Packer, C., Holt, R. D., Hudson, P. J., Lafferty, K. D., and Dobson, A. P. (2003). Keeping the herds healthy and alert: implications of predator control for infectious disease. *Ecology Letters*, 6(9):797–802.
- Panigrahi, A. and Azad, I. S. (2007). Microbial intervention for better fish health in aquaculture: the Indian scenario. *Fish physiology and biochemistry*, 33(4):429–440.
- Pearl, J. (1998). Graphs, Causality, and Structural Equation Models. *Sociological Methods & Research*, R-253:226–284.
- Pearl, J. (2000a). Causality. *New York: Cambridge*.
- Pearl, J. (2000b). Causality. *New York: Cambridge*.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical ..., pages 559–572*.
- Peña, D. and Rodríguez, J. (2003). Descriptive measures of multivariate scatter and linear dependence. *Journal of Multivariate Analysis*, 85(2):361–374.

- Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., Kessing, B., Pontius, J., Roelke, M., Rumpler, Y., Schneider, M. P. C., Silva, A., O'Brien, S. J., and Pecon-Slattery, J. (2011). A Molecular Phylogeny of Living Primates. *PLoS Genet*, 7(3):e1001342.
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. a., and McCarthy, M. a. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, 5(5):397–406.
- Poorter, L., Bongers, L., and Bongers, F. (2006). Architecture of 54 moist-forest tree species: traits, trade-offs, and functional groups. *Ecology*, 87(5):1289–1301.
- Preston, D. L., Henderson, J. S., and Johnson, P. T. (2012). Community ecology of invasions: direct and indirect effects of multiple invasive species on aquatic communities. *Ecology*, 93(6):1254–1261.
- Promislow, D. E. L. and Harvey, P. H. (1990). Living fast and dying young: A comparative analysis of life-history variation among mammals. *Journal of Zoology*, 220(3):417–437.
- Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, a. M., Solan, M., van der Gast, C. J., and Young, J. P. W. (2007). The role of ecological theory in microbial ecology. *Nature reviews. Microbiology*, 5(5):384–392.
- Pullin, A. S. and Stewart, G. B. (2006). Guidelines for systematic review in conservation and environmental management. *Conservation biology : the journal of the Society for Conservation Biology*, 20(6):1647–56.
- Puoti, M., Airoldi, M., Bruno, R., Zanini, B., Spinetti, A., Pezzoli, C., Patroni, A., Castelli, F., Sacchi, P., Filice, G., and Carosi, G. (2002). Hepatitis B virus co-infection in human immunodeficiency virus infected subjects. *AIDS Reviews*, 4:27–35.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Randolph, S. E. and Dobson, a. D. M. (2012). Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology*, pages 1–17.
- Reeder, N. M. M., Pessier, A. P., and Vredenburg, V. T. (2012). A reservoir species for the emerging amphibian pathogen Batrachochytrium dendrobatidis thrives in a landscape decimated by disease. *PLoS ONE*, 7(3):1–7.
- Reich, B. J. and Bandyopadhyay, D. (2010). A latent factor model for spatial data with informative missingness. *Annals of Applied Statistics*, 4(1):439–459.
- Richgels, K. L. D., Hoverman, J. T., and Johnson, P. T. J. (2013). Evaluating the role of regional and local processes in structuring a larval trematode metacommunity of *Helisoma trivolvis*. *Ecography*, 36(7):854–863.
- Ricklefs, R. E. and Wikelski, M. (2002). The physiology/life-history nexus. *Trends in Ecology & Evolution*, 17(10):462–468.

- Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004). Metagenomics: genomic analysis of microbial communities. *Annual Review of Genetics*, 38(1):525–552.
- Robins, J. D. and Vollmar, J. E. (2002). Livestock grazing and vernal pools. In *Wildlife and Rare Plant Ecology of Eastern Merced County's Vernal Pool Grasslands*, pages 401–430.
- Roche, B., Dobson, A. P., Guégan, J.-F., and Rohani, P. (2012). Linking community and disease ecology: the impact of biodiversity on pathogen transmission. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1604):2807–13.
- Roche, B., Rohani, P., Dobson, A. P., and Guégan, J.-F. (2013). The impact of community organization on vector-borne pathogens. *The American naturalist*, 181(1):1–11.
- Roff, D. A. (2001). *Life History Evolution*. Sinauer Associates, Sunderland, MA.
- Rohde, C. L., Bartolini, V., and Jones, N. (2009). The use of probiotics in the prevention and treatment of antibiotic-associated diarrhea with special interest in Clostridium difficile-associated diarrhea. *Nutrition in Clinical Practice*, 24(1):33–40.
- Rouse, J. D., Bishop, C. A., and Struger, J. (1999). Nitrogen pollution: an assessment of its threat to amphibian survival. *Environmental Health Perspectives*, 107(10):799–803.
- Rowe, C. and Dunson, W. (1995). Impacts of hydroperiod on growth and survival of larval amphibians in temporary ponds of central Pennsylvania, USA. *Oecologia*, 102(4):397–403.
- Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(March):108–115.
- Royle, J. A. and Kéry, M. (2007). A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, 88(7):1813–1823.
- Royle, J. A. and Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4):835–841.
- Royle, J. A., Nichols, J. D., Kéry, M., and Ranta, E. (2005). Modelling occurrence and abundance of species when detection is imperfect. *Oikos*, 110(2):353–359.
- Rudolf, V. H. W. and Antonovics, J. (2005). Species coexistence and pathogens with frequency-dependent transmission. *The American naturalist*, 166(1):112–8.
- Rykiel, E. J. (1996). Testing ecological models: The meaning of validation. *Ecological Modelling*, 90(3):229–244.
- Saether, B.-E., Grøtan, V., Engen, S., Noble, D. G., and Freckleton, R. P. (2011). Rarity, life history and scaling of the dynamics in time and space of British birds. *The Journal of animal ecology*, 80(1):215–24.
- Salkeld, D. J., Padgett, K. a., and Jones, J. H. (2013). A meta-analysis suggesting that the relationship between biodiversity and risk of zoonotic pathogen transmission is idiosyncratic. *Ecology letters*, 16(5):679–86.
- Schmutz, A. C., Gray, M. J., Burton, E. C., and Miller, D. L. (2008). Impacts of cattle on amphibian larvae and the aquatic environment. *Freshwater Biology*, 53(12):2613–2625.

- Schnyder, M., Stärk, K. D. C., Vanzetti, T., Salman, M. D., Thor, B., Schleiss, W., and Griot, C. (2002). Epidemiology and control of an outbreak of classical swine fever in wild boar in Switzerland. *The Veterinary record*, 150(4):102–9.
- Schrag, S. J. and Wiener, P. (1995). Emerging infectious disease: what are the relative roles of ecology and evolution? *Trends in Ecology & Evolution*, 10(8):319–324.
- Schwanz, L. E. (2008). Chronic parasitic infection alters reproductive output in deer mice. *Behavioral Ecology and Sociobiology*, 62(8):1351–1358.
- Shapiro, a. (1985). Identifiability of factor analysis: some results and open problems. *Linear Algebra and its Applications*, 70:1–7.
- Shaw, J. D., Wilson, J. R. U., and Richardson, D. M. (2010). Initiating dialogue between scientists and managers of biological invasions. *Biological Invasions*, pages 4077–4083.
- Shipley, B. (2002). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge University Press.
- Shipley, B. (2013). The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology*, 94(3):560–564.
- Shu, Q., Qu, F., and Gill, H. S. (2001). Probiotic treatment using *Bifidobacterium lactis* HN019 reduces weanling diarrhea associated with rotavirus and *Escherichia coli* infection in a piglet model. *Journal of pediatric gastroenterology and nutrition*, 33(2):171–177.
- Stan Development Team (2014a). RStan: the R interface to Stan, Version 2.5.0.
- Stan Development Team (2014b). Stan: A C++ Library for Probability and Sampling, Version 2.5.0.
- Stan Development Team (2014c). *Stan Modeling Language Users Guide and Reference Manual*, Version 2.5.0.
- Stearns, S. C. (1992). *The Evolution of Life Histories*. Oxford University Press, USA.
- Stebbins, R. C. (2003). *A Field Guide to Western Reptiles and Amphibians*. Peterson field guide series. Houghton Mifflin.
- Stine, R. A. (1995). Graphical Interpretation of Variance Inflation Factors. *The American Statistician*, 49(1):53–56.
- Suzán, G., García-Peña, G. E., Castro-Arellano, I., Rico, O., Rubio, A. V., Tolsá, M. J., Roche, B., Hosseini, P. R., Rizzoli, A., Murray, K. A., Zambrana-Torrelio, C., Vittecoq, M., Bailly, X., Aguirre, A. A., Daszak, P., Prieur-Richard, A. H., Mills, J. N., and Guégan, J. F. (2015). Metacommunity and phylogenetic structure determine wildlife and zoonotic infectious disease patterns in time and space. *Ecology and Evolution*, 5(4):865–873.
- Szuroczki, D. and Richardson, J. M. L. (2011). Palatability of the larvae of three species of *Lithobates*. *Herpetologica*, 67(3):213–221.
- Telfer, S., Lambin, X., Birtles, R., Beldomenico, P., Burthe, S., Paterson, S., and Begon, M. (2010). Species interactions in a parasite community drive infection risk in a wildlife population. *Science*, 330(6001):243–246.

- Tsao, J. I., Wootton, J. T., Bunikis, J., Luna, M. G., Fish, D., and Barbour, A. G. (2004). An ecological approach to preventing human infection: vaccinating wild mouse reservoirs intervenes in the Lyme disease cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52):18159–64.
- Vehtari, A., Gelman, A., and Gabry, J. (2015). Efficient implementation of leave-one-out cross-validation and WAIC for evaluating fitted Bayesian models.
- Ver Hoef, J. M. and Boveng, P. L. (2007). Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology*, 88(11):2766–2772.
- Vié, J.-C., Hilton-Taylor, C., and Stuart, S. N. (2009). *Wildlife in a changing world: an analysis of the 2008 IUCN Red List of threatened species*. IUCN.
- Walters, B. (1975). Studies of interspecific predation within an amphibian community. *Journal of Herpetology*, 9:267–279.
- Warton, D. I., Blanchet, F. G., Hara, R. B. O., Ovaskainen, O., Taskinen, S., Walker, S. C., and Hui, F. K. C. (2015). So many variables: joint modeling in community ecology. *Trends in Ecology & Evolution*.
- Wasserberg, G., Osnas, E. E., Rolley, R. E., and Samuel, M. D. (2009). Host culling as an adaptive management tool for chronic wasting disease in white-tailed deer: a modelling study. *The Journal of applied ecology*, 46(2):457–466.
- Webb, C. T., Hoeting, J. A., Ames, G. M., Pyne, M. I., and LeRoy Poff, N. (2010). A structured and dynamic framework to advance traits-based theory and prediction in ecology. *Ecology Letters*, 13(3):267–283.
- Welsh, H. H., Pope, K. L., and Boiano, D. (2006). Sub-alpine amphibian distributions related to species palatability to non-native salmonids in the Klamath mountains of northern California. *Diversity and Distributions*, 12(3):298–309.
- West, M. (2003). Bayesian factor regression models in the large p, small n paradigm. *Bayesian Statistics 7 - Proceedings of the Seventh Valencia International Meeting*, pages 723–732.
- Westoby, M., Falster, D. S., Moles, A. T., Vesk, P. A., and Wright, I. J. (2002). Plant ecological strategies: some leading dimensions of variation between species. *Annual Review of Ecology and Systematics*, 33(1):125–159.
- Wild, M. A., Hobbs, N. T., Graham, M. S., and Miller, M. W. (2011). The role of predation in disease control: a comparison of selective and nonselective removal on prion disease dynamics in deer. *Journal of Wildlife Diseases*, 47(1):78–93.
- Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L., Lenoir, J., Damgaard, C. F., Dormann, C. F., Forchhammer, M. C., Grytnes, J. A., Guisan, A., Heikkinen, R. K., Høye, T. T., Kühn, I., Luoto, M., Maiorano, L., Nilsson, M. C., Normand, S., Öckinger, E., Schmidt, N. M., Termansen, M., Timmermann, A., Wardle, D. A., Aastrup, P., and Svenning, J. C. (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews*, 88(1):15–30.

- Wobeser, G. (2002). Disease management strategies for wildlife. 21(1):159–178.
- Wobeser, G. A. (2013). Investigation and management of disease in wild animals. Springer Science & Business Media.
- Wood, C. L. and Lafferty, K. D. (2012). Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. Trends in ecology & evolution, pages 1–9.
- Wood, S. N. and Thomas, M. B. (1996). Space, time and persistence of virulent pathogens. Proceedings. Biological sciences / The Royal Society, 263(1371):673–80.
- Woodroffe, R., Donnelly, C. a., Jenkins, H. E., Johnston, W. T., Cox, D. R., Bourne, F. J., Cheeseman, C. L., Delahay, R. J., Clifton-Hadley, R. S., Gettinby, G., Gilks, P., Hewinson, R. G., McInerney, J. P., and Morrison, W. I. (2006). Culling and cattle controls influence tuberculosis risk for badgers. Proceedings of the National Academy of Sciences of the United States of America, 103(40):14713–7.
- Wright, D. H. (1998). A comparative analysis of nested subset patterns of species composition. Oecologia, pages 1–20.
- Young, H., Griffin, R. H., Wood, C. L., and Nunn, C. L. (2013). Does habitat disturbance increase infectious disease risk for primates? Ecology letters, 16(5):656–63.
- Zani, J. L., Da Cruz, F. W., Dos Santos, A. F., and Gil-Turnes, C. (1998). Effect of probiotic CenBiot on the control of diarrhoea and feed efficiency in pigs. Journal of Applied Microbiology, 84(1):68–71.
- Zheng, L., Ives, A. R., Garland, T., Larget, B. R., Yu, Y., and Cao, K. (2009). New multivariate tests for phylogenetic signal and trait correlations applied to ecophysiological phenotypes of nine Manglietia species. Functional Ecology, 23(6):1059–1069.
- Zipkin, E. F., Campbell Grant, E. H., and Fagan, W. F. (2012). Evaluating the predictive abilities of community occupancy models using AUC while accounting for imperfect detection. Ecological Applications, 22(7):1962–1972.