Title
Everything old is new again: robust predictive frameworks for shifting host-pathogen interactions in the face of global change

Author
Carlson, Colin J

Publication Date
2017

Peer reviewed|Thesis/dissertation

Everything Old is New Again: Robust Predictive Frameworks for Shifting Host-Pathogen Interactions in the Face of Global Change

by

Colin J Carlson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, & Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Wayne Marcus Getz, Chair
Professor Erica Bree Rosenblum
Professor Michael Boots
Professor Eva Harris

Fall 2017

Everything Old is New Again: Robust Predictive Frameworks for Shifting Host-Pathogen Interactions in the Face of Global Change

Abstract

Everything Old is New Again: Robust Predictive Frameworks for Shifting Host-Pathogen Interactions in the Face of Global Change

by

Colin J Carlson

Doctor of Philosophy in Environmental Science, Policy, & Management

University of California, Berkeley

Professor Wayne Marcus Getz, Chair

Disease ecology urgently requires powerful predictive tools that anticipate the links between global change and emerging infectious disease. However, the ecological context of emerging disease remains poorly understood, especially given that the majority of parasites in any given ecosystem have no direct impact on human health. This dissertation explores a global change biology approach to host-pathogen interactions, focused on understanding both positive and negative impacts of climate change on parasites and pathogens. Chapter 1 reviews current theory surrounding extinction, including mathematical modeling approaches at scales from population extirpation up through global extinction rates. Community-level approaches to extinction risk estimation are applied in Chapter 2, which includes forecasts for climate-driven range shifts based on the largest macroparasite occurrence dataset yet assembled. Up to a third of parasites could face extinction in a changing climate, especially accounting for co-extinction with hosts. However, we find no evidence that wildlife parasites face better or worse odds of survival (or have different hotspots of diversity) based on their potential to infect humans. The results of this study indicate the hundreds of thousands, or potentially millions, of parasitic species on Earth are likely to be redistributed around the globe in a hard to predict pattern, with unknown effects on wildlife and human health. The same species distribution modeling methods from Chapter 2 are used in Chapter 3 to predict the global distribution of Zika virus, an emerging infection from 2016 with a still largely unresolved eco-epidemiology. The conflict among different models and modeling approaches surrounding Zika's distribution is considered in Chapter 4, by interfacing these models with simulations of potential epidemics in the United States. Overall, this dissertation addresses the idea that in the face of global change, ecologists will play an increasingly important role in predicting shifting landscapes of disease. However, the overwhelming focus on emergence ignores the importance of extinction as a potentially complementary phenomenon within ecosystems; and the varied approaches within ecology, and the short timescale on which ecologists work during current outbreaks, pose a disciplinary problem with no clear answer.

For Cory

We are like ships passing in the night, especially you.
I think you would've liked the parts with the birds.

Contents

List of Figures

List of Tables

Acknowledgments

Thanks first and foremost to Wayne Getz, the world's most patient advisor, for providing me the most intellectually enriching years of my life, pushing me towards rigor and ambition in my research and my personal life, and most of all for taking on a tremendous risk that no one else would when he accepted me as a student. The Getz lab is my favorite place not just to work but in this part of the world, and I can't express how lucky I've felt to be part of it. Thanks also to my amazing committee (Bree Rosenblum, Mike Boots, and Eva Harris), who pushed me to challenge my ideas about disease ecology, helped me transition into my current research interests, and keeping me interested and excited. Thanks especially to Bree, who mentored me in how to reconcile my personal values with my professional life, and to practice science with integrity.

Thanks to the Getz lab current (Eric Dougherty, Dana Seidel, Oliver Muellerklein, Zhongqi Miao, and Whitney Mgbara) and past (Carrie Cizauskas, Wendy Turner, Pauline Kamath, Holly Ganz, Miriam Tsalyuk, and Sadie Ryan) for the endless brainstorming and support throughout every project. Thanks to Giovanni Castaldo, Humza Siddiqui, Savannah Miller, Fred Heath, and Faith de Amaral for dedicated undergrad work in our lab in the last five years.

Thank you to the various members of the Parasite Extinction Research Project (Veronica Bueno, Kevin Burgio, Christopher Clements, Carrie Cizauskas, Nyeema Harris, Anna Phillips) and the extended team (Jorge Doña, Roger Jovani, Sergei Mironov, Heather Proctor, Oliver Muellerklein, Giovanni Castaldo, Sarah Fourby) for four and a half outstanding years of collaboration over five projects.

Thanks to my partners in crime who have made the practice of science worthwhile in the last four years, including Allison Barner, Alex Bond, Kevin Burgio, Fausto Bustos, Tad Dallas, Auriel Fournier, Kyrre Kausrud, Oliver Keyes, Dave Miller, Isabel Ott, Anna Phillips, and Tim Poisot. When I think of you, I think of that Leslie Knope quote: "What makes work worth doing is getting to do it with people you love." The endless mentorship, career advice, and receptiveness while I screamed into the void was unbelievable. Most of all, a special thanks to Kevin, the world's most patient collaborator, my oldest friend in science, and the other half of our murderous pair.

Thanks to the friends who carried me to the finish line emotionally and sometimes physically, including Laura Alexander, Bryan Bach, Allison Barner, Lewis Bartlett, Felix Donovan, Claire Dubin, Courtney Hodrick, Katherine Spack, and Kirsten Verster.

Thank you to Joel Scheraga for teaching me when to be right, and when to be happy. Thank you to Carl Schlichting for teaching me how to be a scientist in the first place, and for shaping the adult I am today.

Finally, thanks to my family (especially my grandma Carole and my other grandma Carole) for supporting me through the scariest parts; and thanks to my three absolute best friends in the world (Zach Brown, Fausto Bustos, and Emily Kearney), for not giving up on my vision for myself, and helping me get to the end.

Fausto, you've filled my life with laughter and pie, and given me the chance to be the crazy next-door neighbor I've always wanted to be. You're the smartest guy I know, and talking to you about epidemiology, revising the Six Blind Scientists paper at two in the morning, or just listening to you talk about bootstrapping, have been the highlights of this Ph.D. Thank you for keeping me laughing.

Emily, you're the first friend I made in Berkeley, and still my favorite person in California. I can't tell you what it means to have you in my life. I believe in you more than I believe in anyone, and I know you're going to do absolutely amazing things once you figure out what makes you happy. Thank you for keeping me sane.

And finally, to my brother Zach: you remain the most humbling person I have ever met in my life. Your compassion and dedication to improving the world are my highest inspiration, and after all these years you're still my closest family because of it. I know that wherever you are, I can always come home to you, and I can't wait to do that at the end of this. Thank you for pushing me to be the most I can be, and for keeping me kind.

I love you all so much. I spent my last year trying to figure out who I was trying to make proud, and then while I was writing this, I realized it was the three of you.

Chapter 1

Preface: It's the end of the world (as we know it)

In the face of environmental change, all species have a fundamental set of possible responses: acclimate and adapt, move, or face extinction. In the context of disease ecology, the interplay between hosts' and pathogens' basic responses can produce incredibly complex and unexpected patterns of change, with dramatic consequences for human health and the environment. An emerging field of experimental, modeling, and theoretical work centers around the challenge of predicting these patterns and processes, ideally with downstream benefits for both biodiversity conservation and public health projects.

But, by and large, the majority of that work centers around emerging or immediate threats to human health. In the context of climate change research, the majority of work in predicted hotspots of parasitic biodiversity focuses on the responses of pathogens like malaria, dengue fever, and leishmaniasis [1], and comparatively less is understood about wildlife parasites with no human mortality or morbidity. The problem, of course, is one of scale. A quick search of the London Natural History Museum's Host-Parasite Database [2], the most comprehensive database of helminth parasites in existence, finds fewer than 700 human parasites among roughly 20,000 known species of vertebrate helminths. Those are only a subset of the estimated 75,000–300,000 species of vertebrate helminth alone [3]; invertebrate parasites are several orders of magnitude more diverse, with a recently estimated 81 million nematode parasites of arthropods alone. [4] By most accounting methods, the majority of species on Earth are parasites, and almost all of them are undiscovered or undescribed, making them all but impossible to include in global change research.

Any framework for global change biology that excludes parasites is therefore, by definition, missing the majority of species in an ecosystem–and consequently, the ecosystem-level dynamics that emerge from parasitism. For example, in many ecosystems, parasites are responsible for the majority of food web links [3], though they are frequently omitted from food web studies. Parasites can be the majority of biomass in an ecosystem, and their ability to regulate host populations and host behavior have tremendous consequences for ecosys-

tem stability. [5] In some cases, parasites may be responsible for maintaining key species interactions that ultimately benefit endangered species; in one particularly unusual example, behavior-altered crickets parasitized by nematomorph worms jump into streams, and make up the majority of the diet of endangered Japanese trout. [6, 7] The downstream effects of climate change on host-parasite interactions can be expected to disrupt all of these ecosystem roles and services; but almost nothing is known, especially proportional to the diversity of parasites on Earth.

Many parasites are likely to adapt to changing climates, or already have plastic responses to environmental change that will buffer them (and their host associations) from the impacts of global change; some, even, will become more virulent or more prevalent within their current ranges. But other species are likely to face extinction, and the scale of that phenomenon is poorly understood (with even less attention to its potential adverse impact on ecosystems). At the last major estimate, 3-5% of parasites could be estimated to face coextinction alongside vulnerable hosts [3]; but nothing is known about parasite extinction risk from climate change, despite a significant corresponding body of literature surrounding extinction risk of free-living species. [8] Predicting the impacts of ecological change on parasite-inclusive networks proves challenging without this key information; and little theory has been developed that explores the possible relationship between parasite extinction and disease emergence.

In this dissertation, I explore the predictive crisis of disease and global change from both sides (parasite extinction and disease emergence). The first half reviews the extinction side of the problem. Chapter 1, a book chapter prepared for *The Mathematics of Planet Earth*, reviews the existing body of theoretical ecology considering extinction at all scales, from populations up to the planetary level. Chapter 2 applies this body of theory, especially extinction risk estimation using the species area relationship, to quantify the potential negative impacts of climate change on parasite biodiversity. To achieve this, I led a team of 17 researchers in 8 countries in an effort to compile what has become the largest operational parasite occurrence datasets. That data was used with species distribution models [9] to project the distribution of nearly 500 species in current and future climates, and predict habitat loss. Overall, parasite extinction rates could reach levels as high as 30% if climate change and host extinctions act synergistically. But perhaps more interestingly, species that do not go extinct are also predicted to shift their distributions substantially in the next half-century. The aggregation of parasites in new ecosystems, and the loss of native parasite diversity, could create dramatic and unexpected opportunities for disease emergence in wildlife and potentially humans–especially if parasite diversity can buffer against emergence events. [10]

In the second half of this dissertation, I explore the problems associated with using the same method (species distribution modeling, also termed ecological niche modeling) to predict disease emergence. Chapters 3 and 4 focus on Zika virus, a pathogen first described in 1947 that has only recently emerged as a global crisis. Chapter 3 presents what, when first published, was the first ecological niche model for Zika virus. Using a set of basic climatic predictors, models were generated not only for Zika, but for the closely-related dengue fever and three of the *Aedes* mosquitoes that vector these viruses. Results indicated

ZIka virus could be expected to be more strictly tropical than dengue fever, even in the face of climate change, which expands the range of both. However, in the time since that study, two alternate models have been published, with substantial differences among them. [11, 12] Chapter 4 addresses the disagreements between published models, and explores the downstream impacts of those differences on epidemiological simulations, and shows the unacceptable margin of uncertainty that emerges from a lack of *post hoc* consensus building in the disease ecology literature. Chapter 4 concludes by presenting a global, seasonal consensus model for areas of Zika virus transmission risk, a basic but important first step in advancing disease ecologists' capacity to contribute to public health.

The links between parasite ecology and disease emergence are still somewhat underexplored. The macroecology of infectious disease is still a comparatively new field, and community ecology-based approaches to predicting disease emergence are still comparatively underdeveloped. [13, 14] However, as this body of work highlights, global change is a cohesive process with many disparate aspects, and the interactions among them (e.g., feedbacks between parasite and host extinctions and disease emergence) are hard to capture but may be the majority of the changes that predictive approaches aim to anticipate. Given the staggering diversity of parasitic life on Earth, I suggest that statistical and machine-learning approaches are desperately needed that build on this work, and learn from existing patterns to accurately forecast the process of disease emergence without a comprehensive understanding of underlying patterns of biodiversity. Such tools will only become increasingly important as climate change redistributes the parasites and pathogens that constitute most of life on this planet.

Chapter 2

The mathematics of extinction across scales: from populations to the biosphere

Colin J. Carlson Kevin Burgio Tad Dallas Wayne Getz

2.1 Abstract

The sixth mass extinction poses an unparalleled quantitative challenge to conservation biologists. Mathematicians and ecologists alike face the problem of developing models that can scale predictions of extinction rates from populations to the level of a species, or even to an entire ecosystem. We review some of the most basic stochastic and analytical methods of calculating extinction risk at different scales, including population viability analysis, stochastic metapopulation occupancy models, and the species area relationship. We also consider two major extensions of theory: the possibility of evolutionary rescue from extinction in a changing environment, and the posthumous assignment of an extinction date from sighting records. In the case of the latter, we provide a new example using data on Spix's macaw (*Cyanopsitta spixii*), the "rarest bird in the world," to demonstrate the challenges associated with extinction date research.

> It's easy to think that as a result of the extinction of the dodo, we are now sadder and wiser, but there's a lot of evidence to suggest that we are merely sadder and better informed.
>
> – Douglas Adams, *Last Chance to See*

2.2 Introduction

Most species, like most living organisms on Earth, have a finite lifespan. From the origin of a species onward, every species changes and adapts to its environment. Some species exist longer than others, but all eventually face extinction (or, are replaced by their descendants through evolution). Currently, there are approximately 8.7 million eukaryote species alone. But in the history of Earth, it is estimated that there have been a daunting 4 billion species altogether, and at least 99 percent of them are now gone. [15]

How long can a species exist? Of the species currently on Earth, some are deeply embedded in the geological record and have changed very little over the span of million years, such as coelacanths or ginkgo trees. Most species persist for a few millions of years or more, and in periods of environmental stability, extinctions typically occur at a low and steady baseline rate. But at various points in the history of the Earth, extinction rates have suddenly accelerated for brief and eventful periods that biologists term *mass extinction events*. In 1982, based on the marine fossil record, David Raup and Jack Sepkoski suggested that five of these mass extinctions have occurred over the past half billion years. [16] In all five, more than half of all contemporary species disappeared [17], each sufficiently drastic to be considered the end of a geological era: the Ordovician 444 million years ago (*mya*), Devonian 375 mya, Permian 251 mya, Triassic 200 mya and Cretaceous 66 mya.

But in recent years, ecologists have reached the consensus that the biosphere is currently experiencing, or at the very least entering, the sixth mass extinction. [18] Unlike the previous five, which were caused by planetary catastrophes and other changes in the abiotic environment, the sixth mass extinction is the undeniable product of human activities. While anthropogenic climate change is one of the most significant contributors, a number of other factors have recently exacerbated extinction rates, including habitat loss and fragmentation, biological invasions, urbanization, over-harvesting, pollution, pests, and emerging diseases.

How does the sixth mass extinction scale up against the last five? The number of extinctions alone is an unhelpful metric, as species richness changes over time. A more convenient unit of measurement commonly used by scientists is the number of *extinctions per millions of species-years* (E/MSY). From a landmark study by Gerardo Ceballos and colleagues, we know that in the geological record, vertebrates normally go extinct at a rate of 2 E/MSY in the periods in-between mass extinctions. But since 1900, that rate is an astounding 53 times higher. [19] One study has suggested that the sixth mass extinction is comparable to other mass extinctions in E/MSY rates, meaning that with enough time, the geological definition of a mass extinction (three quarters extinction) could be achieved in hundreds to thousands of years. [20] Or, to consider another framing : a 1970 study estimated that at a baseline, one species goes extinct per year [21], while just a decade later that estimate was revised to one species per hour. [22] Plants, insects, and even micro-organisms all face similarly catastrophic threats; and these across-the-board losses of biodiversity pose a threat to human survival that some argue could even threaten our own species with extinction.

The crisis of extinction is, for scientists, a crisis of prediction. While extinction is a natural part of ecosystem processes and of the history of the planet, the job of conservation biologists

is to protect species that would otherwise be brought to an untimely and avoidable end. To do that, conservationists must sort and prioritize the 8.7 million eukaryotes (and, even, some prokaryotes) to assess which species face the greatest threat—and which can, and cannot, be saved by human intervention. Assessment is easiest at the finest scales: by marking and tracking all the individuals in a region, a population ecologist can make a statistically-informed estimate of the probability of imminent extinction. But above the population level, assessment is much more challenging, requiring sophisticated (and complicated) meta-population models that are typically data-intensive. If a species is rare enough and the data are "noisy," its extinction may seem uncertain even after the fact; but mathematical models can help assign a probability to the rediscovery of a species once thought extinct, and resolve when (and even why) a species has disappeared long after it is gone. Above the level of a single species, measuring extinction is an altogether different problem, requiring a different type of model to explain how biodiversity arises and is maintained over time.

Each of these modeling approaches represents a different aspect of a connected problem, and we deal with each in turn in this chapter. The models we present are seminal and well-known, but extinction risk modeling is a dynamic and rapidly-growing field. Consequently, these models only present a handful of the many different approaches that link different temporal and spatial scales of extinction together.

2.3 The Population Scale

Even though many make a terminological distinction between *extinction* (the loss of a species) and *extirpation* (the eradication of a population), extinction is still fundamentally a process that begins at the population scale. With the exception of sudden, unexpected catastrophes, extinction at the population scale is almost always the product of either a declining population or of stochastic variations in an already-small population, both of which follow mathematical rules that can be used to quantify extinction risk. Perhaps the most significant body of theory about population extinction deals with the estimation of a population's *mean time to extinction* (MTE, typically T_E in mathematical notation), an important quantity to both theoretical ecologists and conservation practitioners. For both theoretical and applied approaches to extinction, understanding the uncertainty around T_E requires an understanding of the shape of the extinction time distribution, including developing and testing demographic theory that accurately captures both the central tendencies [23] and the long tail [24] of empirical extinction times. We begin by reviewing some of the basic population-scale approaches that scale up to ecosystem-level theory of extinction.

Stochasticity and the Timing of Extinction

In the most basic terms, a population declining at a steady rate will eventually become extinct; the simplest deterministic equation governing the size of a population N, as it grows over time t (generally measure in units of either years or generations) is given by

$$\frac{dN}{dt} = rN \tag{2.1}$$

where if r is positive the population is growing, while if r is negative, the population heads rapidly towards extinction. A slightly more complex model that captures the phenomenological capping of the growth of a population at a population ceiling termed K is:

$$\frac{dN}{dt} = \begin{cases} rN & \text{if } 1 < N < K \\ 0 & \text{if } N = K \end{cases} \tag{2.2}$$

While K is often called a carrying capacity, this is perhaps misleading, as in this context it only introduces density dependence when $N = K$, and not before. Eqns. 2.1 and 2.2 both imply that if $r < 0$, $\ln(N)$ declines linearly with slope r. For shrinking populations (i.e., $r < 0$) these equations imply that the mean time to extinction (T_E) can be derived analytically as the amount of time before the population reaches one individual (i.e. $N = 1$ at $t = T_E$):

$$T_E(N_0) = -\ln(N_0)/r \tag{2.3}$$

Consequently, for a given population with a fixed r the maximum achievable extinction time given a starting stable population size would be

$$\max(T_E) = -\ln(K)/r \tag{2.4}$$

However, in this model if $r > 1$, the population never reaches extinction and simply grows forever.

Deterministic models only tell a part of the story. In the history of conservation biology, two paradigms emerged that separately explain the process of population extinctions. The *declining population paradigm* explains that populations shrink and vanish due to a combination of internal and external failures, and suggests that the key to conserving populations is to identify and prevent those failures. In contrast, the *small population paradigm* is rooted in ideas of stochasticity, suggesting that even without factors like environmental degradation or disease, smaller, more fragmented populations simply face higher extinction risk due to stochastic population processes. [25] For one thing, stochasticity produces populations with log-normally distributed sizes (i.e. most populations are comparatively small relative to a few larger ones). The underlying reason for this can be traced back to Jensen's inequality, which suggests the expected value of a convex function applied to a random variable x is greater than, or equal to, that function applied to the expected value of the random variable (below, $E[\cdot]$ is the expectation operator):

$$E[f(x)] \geq f(E[x])$$

Applied to stochastic population growth, if r is stochastic, the expectation of r will always be greater than the expected real growth rate of the population [26]:

$$E[r] > E[(N_t/N_0)^{1/t}]$$

Iterating these lower growth rates over an infinite amount of time, populations that are growing randomly with $\bar{r} \leq 1$ (i.e. less than exponential growth) all tend eventually to extinction.

In general, r can be decomposed into two component processes; births and deaths. In their foundational work on the ecology of invasion and extinction—*The Theory of Island Biogeography*—Robert MacArthur and E.O. Wilson proposed a simple model with discrete per-capita birth and death rates, λ and μ respectively. With $\lambda + \mu$ changes expected per time step, the estimated time until a single change (birth or death) is given $1/(\lambda + \mu)$. Thus the time to extinction for a population of x individuals, $T_E(x)$ can be intuitively understood (with a more detailed derivation in MacArthur and Wilson) as: i) the expected time for one change to occur (birth or death); plus ii) the probability the change is a birth (i.e., $\frac{\lambda}{\lambda+\mu}$) multiplied by the time to extinction if the population is of size $x-1$; plus iii) the probability the change is a death (i.e., $\frac{\lambda}{\lambda+\mu}$) multiplied by the time to extinction if the population is of size $x+1$.

This reasoning produces the relationship:

$$T_E(x) = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}T_E(x+1) + \frac{\mu}{\lambda+\mu}T_E(x-1)$$

This simple but elegant relationship can be used to produce an expression for $T_E(K)$ using the method of induction; in particular, $T_E(x)$ can be expressed as a function of $T_E(1)$, noting that $T_E(0) = 0$. To do this, MacArthur and Wilson add a population ceiling K as before, and consider two cases of density dependence. If births are density dependent, then (using the notation $\tilde{\lambda}(x)$ to distinguish between the function $\tilde{\lambda}$ and the constant λ and similarly for μ)

$$\tilde{\lambda}(x) = \begin{cases} \lambda x & \text{if } X < K \\ 0 & \text{if } X \geq K \end{cases}$$

$$\tilde{\mu}(x) = \mu x$$

and (through and inductive procedure not shown here) the time to extinction is

$$T_E(K) = \frac{\lambda}{\lambda-\mu}T_E(1) + \frac{\lambda}{\mu(K+1)(\lambda-\mu)} - \frac{1}{\lambda-\mu}\sum_{i=1}^{K}\frac{1}{i}$$

where

$$T_E(K) = \frac{1}{\mu}(K+1)$$

and

$$T_E(1) = \sum_{i=1}^{K} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i\lambda} + \left(\frac{\lambda}{\mu}\right)^K \frac{1}{\mu(K+1)}$$

In contrast, if and when deaths are density dependent,

$$\tilde{\lambda}(x) = \lambda x$$

$$\tilde{\mu}(x) = \begin{cases} \mu x & \text{if } X < K \\ 0 & \text{if } X \geq K \end{cases}$$

and the time to extinction is

$$T_E(K) = \frac{\lambda}{\lambda - \mu} T_E(1) - \frac{1}{\lambda - \mu} \sum_{i=1}^{K} \frac{1}{i}$$

In this scenario, $T_E(K) = T_E(K+1)$; with some induction (not shown here), $T_E(1)$ can also be expressed as

$$T_E(1) = \sum_{i=1}^{K} \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i\lambda}$$

This provides an explicit method for calculating $T_E(K)$, the maximum achievable time to extinction with these rates. MacArthur and Wilson made a handful of key observations about the behavior of these functions as they relate both to island biogeography and to the population process of extinction. First, $T_E(1)$ can be surprisingly large if $\lambda > \mu$, meaning that a net tendency for growth has incredibly long times before extinction, even with stochasticity. Second, if populations start with a single propagule (as their work is framed in the context of island colonists), roughly μ/λ go extinct almost immediately while roughly $(\lambda - \mu)/\lambda$ grow to K and take $T_E(K)$ years to go extinct. (This means that even though density dependence is not introduced until $N = K$, the effects of the population ceiling are still emergent on the dynamics of the whole system.) Third, "established populations" ($N = K$) have a readily calculated extinction time:

$$T_E(K) \approx \frac{\lambda}{\lambda + \mu} T_E(1) = \frac{\lambda}{r} T_E(1)$$

When $\lambda > \mu$, the time to extinction scales exponentially with the population ceiling, and does so at a hyperbolically accelerating rate with r. In short, bounded random birth-death processes still approach extinction, but do so incredibly slowly if populations tend towards growth.

To more explicitly determine time to extinction in an exponentially growing population, consider a population subject to simple *Weiner process* type stochastic fluctuations $W(t)$.

[27, 28] Specifically, if dW represents the derivative of $W(t)$ such that $W(0) = 0$, then $W(t)$ is normally distributed around 0 such that

$$W(t) \sim \mathcal{N}(0, t) \tag{2.5}$$

and the model is written as

$$dN = rNdt + \int NdW$$

This stochastic differential equation implies that for moderate population sizes, where environmental stochasticity prevails over demographic stochasticity (discussed more fully in the next section), then for levels of infinitesimal environmental variance σ^2, the expected change in log population size $X = \log(N)$ over a small interval $[t, t + h]$ is [27, 28]

$$E[X(t + h) - X(t)] \sim \mathcal{N}\left(\mu h, \sigma^2 h\right) \quad \text{where} \quad \mu = r - \sigma^2/2 \tag{2.6}$$

Solving the stochastic differential equation provides a distribution for X at time t:

$$g(X) = \frac{1}{\sigma\sqrt{2\pi t}}\left(1 - \exp\left(\frac{-2XX_0}{\sigma^2 t}\right)\right)\exp\left(-\frac{(X - X_0 - \mu T)^2}{2\sigma^2 T}\right)$$

Consequently, the distribution of the time to extinction (a population size of $X = 0$, i.e. $N = 1$) is

$$f(T) = \frac{X_0}{\sigma\sqrt{2\pi t}}\exp\left(-\frac{(X_0 + \mu T)^2}{2\sigma^2 T}\right)$$

If $\mu \leq 1$, this integrates to zero; otherwise, it integrates to $1 - \exp\left(-2\mu X_0/\sigma^2\right)$. Combining these expressions,

$$P(T < T_E) = \int_0^\infty g(X)dX = \int_1^\infty f(T)dT$$

gives the probability that the population persists to time T without going extinct.

In reality, populations show a combination of deterministic and stochastic behavior over time, and their extinction is a product of both. In the late 1980s, the field of *population viability analysis* (PVA) emerged from the need to find appropriate analytical and simulation methods for predicting population persistence over time. According to one history of PVA, Mark Shaffer's work on grizzly bears in Yellowstone helped birth the field through two important developments, which we break down in turn below. [29]

Demographic and Environmental Stochasticity

Shaffer's first major contribution was the use of extinction risk simulations that account for—and differentiate between—two major kinds of stochasticity. *Demographic stochasticity* is defined at the scale of the individual and occurs through random variation in demography and reproduction, while *environmental stochasticity* occurs at a synchronized scale for an entire

population (e.g., a bad year may change vital rates uniformly for all individuals in a population). While the impact of environmental stochasticity is ultimately scale-independent, larger populations become less sensitive to demographic stochasticity as they grow. This is due to the integer-based nature of birth-death processes, where populations made up of fewer individuals will suffer a disproportionate effect from a birth or death event.

Demographic and environmental stochasticity have measurably different effects on T_E in basic population models. A simple modeling framework distinguishing between them was laid out in a 1993 paper by Russell Lande. [30] That framework begins again with Eq. 2, except that we now regard r as an explicit function of time $r(t)$ with a mean $\bar{r}$. In the case of demographic stochasticity, individual variations have no temporal autocorrelation and at the population scale,

$$r(t) \sim \mathcal{N}(\bar{r},\ \sigma_d^2/N)$$

where σ_d^2 is the variance of a single individual's fitness per time. As above, the population can be expressed as a diffusion process from the initial population size N_0:

$$\frac{1}{2}\sigma^2(N_0)\frac{d^2T_E}{dN_0^2} + \mu(N_0)\frac{dT_E}{dN_0} = -1$$

The solution of that differential equation for T_E (where extinction happens at $N = 1$) is given as a function of the initial population size:

$$T_E(N_0) = 2\int_1^{N_0} e^{-G(z)} \int_z^K \frac{e^{G(z)}}{\sigma^2(y)} dy dz$$

where

$$G(y) = 2\int_1^y \frac{\mu(N)}{\sigma^2(N)} dN$$

For populations experiencing demographic stochasticity and starting at their carrying capacity, this gives us an expression for extinction time that is perhaps slightly clearer:

$$T_E = \left(\frac{1}{\bar{r}}\int_1^K \frac{e^{2\bar{r}(N-1)N/\sigma_d^2}}{N} dN\right) - \frac{\ln K}{\bar{r}}$$

Thus Lande argues in [30] that when $\bar{r}$ is positive, MTE scales exponentially with carrying capacity, while when $\bar{r}$ is negative it scales logarithmically with carrying capacity (i.e., $T_E \propto \ln(K)$), much like in the deterministic decline given by Eqs. 2.3 & 2.4). In contrast, in the case of environmental stochasticity, the variance acts on the entire population at once (cf. Eq. 2.6):

$$E[\ln N(t)] = \ln N_0 + (\bar{r} - \sigma_e^2/2)t$$

and the mean time to extinction is now given by [30]

$$T_E = \frac{2}{V_e c}\left(\frac{K^c - 1}{c} - \ln K\right)$$

where

$$c = \frac{2\bar{r}}{\sigma_e^2} - 1$$

In the case of environmental stochasticity, if the "long-run growth rate" ($\tilde{r} = \bar{r} - \sigma_e^2/2$) is zero or negative, MTE again scales logarithmically with K. When long-run growth is positive, the dynamic is a bit more complicated:

$$T_E \approx 2K^c/(\sigma_e^2 c^2) \quad \text{if} \quad c \ln K >> 1$$

In this case, the scaling of MTE with K bends up if and only if $\bar{r}/\sigma_e^2 > 1$ (i.e., if and only if the intrinsic growth rate exceeds environmental variation).

Minimum Viable Populations and Effective Population Size

The second major contribution of Shaffer's work was the introduction of the concept of a *minimum viable population* (MVP). In Shaffer's original work, MVP is defined as the smallest possible population for which there is a 95% chance of persistence (a 5% or lower chance of extinction) after 100 years. In their foundational treatment of the minimum viable population concept, Gilpin and Soulé [31] identify four special cases—*extinction vortices*—in which a population is likely to tend below the MVP and towards ultimate extinction.

The first, the *R Vortex*, is perhaps the most obvious: demographic stochasticity (variation in r) reduces populations and increases variation in r, a positive feedback loop of demographic stochasticity directly driving populations to extinction. The *D Vortex* occurs when the same processes—potentially in concert with external forces—produce increased landscape fragmentation (see §3.1.1 for an explanation of D), which not only reduces local population sizes (increasing local extinction rate) but also has subtle effects on population genetic diversity. The final two vortices—the *F Vortex* and *A Vortex*—both concern the genetic and evolutionary trajectories of small stochastic populations. In the first, inbreeding and demographic stochasticity form a feedback cycle, while in the latter, maladaptation is the underlying mechanism of extinction. Both are especially relevant in research surrounding phenomena like climate change, but fully understanding them requires a mathematical language for the genetic behavior of near-extinction populations.

In heavily subdivided populations with low dispersal, increased inbreeding can lead to decreased genetic diversity and the accumulation of deleterious or maladapted alleles that make the total population less viable than its size might indicate. As a consequence, intermediate-sized populations with low genetic diversity can behave, demographically, like small populations. *Effective population size*, or N_e, quantifies that phenomenon, expressing the genetically or reproductively "effective" number of individuals in a population. In some cases, measuring population size with N_e may more readily allow the computation of a meaningful and

predictive MVP, by removing some of the variability between different populations of the same size, and by more accurately capturing the long-term reproductive potential of the available genetic material. (Relatedly, it is worth noting that in one unusual study, it was found that there is no statistical link between species MVP and global conservation status. [32])

A number of different approaches exist for the estimation of N_e. Sewall Wright, who created the concept of effective population size, offered one interpretation based on neighborhoods. In his model, offspring move a distance away from their parent based on a two-dimensional spatial normal distribution with the standard deviation σ. [33] If individuals have a density D, then

$$N_e = 4\pi\sigma^2 D$$

Wright[34] also provides a more commonly invoked method of calculating N_e based on sex structure, using N_m and N_f to respectively denote the number of breeding females and males in the population:

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

In such an approach, a population of all males or all females would have an N_e of 0 (because no new offspring could be produced in the next generation, rendering the population functionally extinct). That method of deriving N_e is still frequently cited in population conservation work to the present day, as small populations tend to stochastically deviate from a 50:50 sex ratio, sometimes severely impacting long-term survival.

A more genetics-based method of calculating N_e comes from the Wright-Fisher model of a two-allele one-locus system, referred to as the *variance effective population size*. [35] In that model, variance between generations $\sigma^2(a)$, for allele A with frequency a, is given by $a(1-a)/2N$, yielding an effective population size of

$$N_e = \frac{a(1-a)}{2\sigma^2}$$

Alternatively, for a locus with a greater degree of polymorphism, or multi-locus microsatellite data, genetic diversity θ and mutation rate μ are related by

$$N_e = \frac{\theta}{4\mu}$$

A more commonly used metric in current literature is *inbreeding effective population size*. To construct that metric, we start by defining population-level measures of heterozygosity. In the simplest Hardy-Weinberg formulation for a two allele system with allele frequencies a and $1-a$, the expected fraction of heterozygote offspring $E(H) = 2a(1-a)$. By counting the real fraction of heterozygotes and comparing, we can measure the assortiveness of mating:

$$f = \frac{E(H) - H}{H}$$

That value f is called the inbreeding coefficient, ranging from 0 to 1; again according to Wright[36], N_e should be calculated such that it satisfies

$$N_e = \frac{1}{2\Delta f}$$

where Δf is the change per generation (in a declining or small population, genetic diversity decreases at a rate determined by the population size and inbreeding).

Returning to the extinction vortex concept with N_e in mind clarifies the genetic component of those extinction processes. While the *D Vortex* reduces N_e as a byproduct of fragmentation (in fact, decreasing neighborhood size), the last two extinction vortices bring N_e below the MVP through specifically genetic modes of extinction. In the *F Vortex*, a positive feedback loop between increased inbreeding (hence f, the inbreeding coefficient) and decreases in effective population size drive a population to extinction over a few generations. A notorious real-world example of such a process might be the near-extinction (or extinction, depending on your species concept) of the Florida panther, a subspecies of *Puma concolor* ultimately rescued through outbreeding with Texas panthers. All things considered, their rescue was both fortuitous and improbable, as the species was assigned a 5% or less chance of avoiding imminent extinction in 1995. [37] Finally, in the *A vortex* (i.e., adaptation), decreased N_e acts as a buffer to the strength of selection acting on phenotypes that are closely paired with environmental variation or change, leading to mismatch between them that reduces both r and N (and N_e) until extinction (a process we cover in much greater detail in §4.1) . Obviously, the four vortices are non-independent processes, and probably often exist in combination in real-world cases.

Population Viability Analysis: Theory and Practice

Population viability analysis is conventionally implemented by modeling the dynamics of different compartmental classes within a population, such as age and sex structure. The foundations of that method date as far back as P. H. Leslie's population analyses in the late 1940s in the framework of discrete matrix models and linear systems theory. Formulations of the Leslie model and the theory behind such models can be found in several expository texts [38, 39], with a brief outline provided here. In the Leslie model, the population is divided into n age classes, where $N_i(t)$ is used to denote the number of individuals in age class i at time t. In each age class, the parameter s_i ($0 < s_i \leq 1$) is used to represent the proportion of individuals aged i that survive to age $i+1$, in which case the variables $N_i(t)$ and $N_{i+1}(t+1)$ are linked by the equation

$$N_{i+1}(t+1) = s_i N_i(t) \tag{2.7}$$

At some point we either terminate this series of equations at age n by assuming that $s_n = 0$ (i.e. no individuals survive beyond age n) or we interpret N_n as the group of individuals in the population aged n and older and use the equation

$$N_n(t+1) = s_{n-1}N_{n-1}(t) + s_nN_n(t) \tag{2.8}$$

to imply that all individuals aged n and older are subject to the survival parameter s_n (i.e., individuals older than age n are indistinguishable from individuals aged n). If we now interpret $N_0(t)$ as all newborn individuals born just after individuals have progressed one age class, then $N_0(t)$ can calculated using the formula

$$N_0(t) = \sum_{i=1}^{n} b_iN_i(t) \tag{2.9}$$

where b_i is the average (expected) number of progeny produced by each individual aged i. In this model we have not differentiated between the sexes; so, for example, if each female aged i is expected to produce 3 young and the population has a 1:1 sex ratio (same number of males to females) then $b_i = 1.5$ for this age class. If we now apply Equation 2.7 for the case $i = 0$, we obtain the equation

$$N_1(t+1) = s_0N_0(t) = s_0\sum_{i=1}^{n} b_iN_i(t) \tag{2.10}$$

Equations 2.7 to 2.10 can be written compactly in matrix notation (a *Leslie matrix*) as

$$\mathbf{N}(t+1) = L\mathbf{N}(t) \tag{2.11}$$

$$\text{where } \mathbf{N} = \begin{pmatrix} N_1 \\ \vdots \\ N_n \end{pmatrix} \text{ and } L = \begin{pmatrix} s_0b_1 & \cdots & s_0b_{n-1} & s_0b_n \\ s_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & s_{n-1} & s_n \end{pmatrix}$$

The matrix L is a nonnegative matrix since all its elements are non-negative, with at least one positive element. Further, if there exists some integer $p > 0$ such that L^p is positive (i.e. all its elements are positive), then it is is known from the Perron-Frobenius Theorem that the matrix L has a dominant positive eigenvalue λ_p (known as the Perron root) and a corresponding eigenvector $\mathbf{v}_p$ whose elements are all positive. These values λ_p and $\mathbf{v}_p$ characterize the long term behavior of $\mathbf{N}$ such that

$$\mathbf{N}(t) \sim (\lambda_p)^t\mathbf{v}_p$$

This equation implies that as t gets very large $\mathbf{N}(t)$ grows like $(\lambda_p)^t$ and the ratio of different age classes matches the ratio of elements of $\mathbf{v}_p$. Thus, if $\lambda_p > (<)1$, $\mathbf{N}(t)$ will grow (decline) geometrically at the rate λ_p and approach the so-called *stable age-distribution*, as

characterized by the ratio of consecutive elements of $\mathbf{v}_p$. In other words, this model predicts that the population will go extinct whenever the largest eigenvalue of L is less than one (i.e., $0 < \lambda_p < 1$). On the other hand, if $\lambda_p > 1$, then we expect density-dependent effects at some point to rein in the unfettered growth by causing survival rates to decline. In particular, if the survival rate s_0 of the youngest age class is the most sensitive of the survival rates to increases in the total biomass density

$$B = \sum_{1}^{n} w_i N_i \tag{2.12}$$

of the population, where $w_i > 0$ is the average weight of an individual in age class i, then we should replace s_0 in Eqn. 2.10 with an expression such as

$$s_0 = \frac{\hat{s}_0}{1 + (B/K_0)^\gamma} \tag{2.13}$$

where $\hat{s}_0$ is the density-independent survival rate, K_0 is the density at which $\hat{s}_0$ is halved, and $\gamma > 1$ is termed the "abruptness" (as it controls the abruptness in the onset of density, approaching a step down function as γ gets large [40]). Similar modifications can be made to the other survival parameters s_i, depending on their sensitivity to changes in population density.

Stochastic equivalents of these deterministic models typically treat the survival rates s_i as probabilities that each individual survives each time period, rather than as the proportion of individuals surviving each time period; and b_i itself is a random variable drawn from an appropriately defined distribution (usually the binomial distribution). Stochastic models of this sort can be made even more complex by adding more population structure (e.g. genetic variability) or increased levels of complexity (e.g. modeling at the metapopulation scale, discussed in §3, or adding underlying environmental variation or other landscape structure). Though MVP or extinction rates might be difficult to calculate analytically for models of this level of complexity, repeated simulation can easily allow empirical derivation of these properties of a system [41], and is perhaps the most widespread practice in existence for estimating population extinction risk in conservation research. An example using an interactive web app [42] is show in Figure 1.1.

Is population viability analysis the perfect tool for studying extinction? PVA is currently the gold standard for most applied conservation research, both by virtue of being an all-encompassing term for quantitative extinction risk modeling at the population scale, and the absence of any suitable alternative. But PVA, like any quantitative tool, is tremendously sensitive to assumptions, parameterization, and data availability. Imprecise parameterization, from noisy data or tenuous assumptions, proportionally reduces the precision of PVA, to a degree that may be hard to characterize; it is consequently important to report uncertainty from PVA estimates [43]. Similarly, given the challenges of developing an accurate and precise model, it has been widely agreed that PVA should be treated as more of a relative or comparative tool (for instance, between different management or conservation scenarios),

and authors should refrain from treating minimum viable population or extinction time estimates as absolute, precise estimates. [44, 45, 46] Despite this, many managers still use PVA as an absolute estimate of extinction risk, a pervasive problem with no clear solution.

Case Study: PVA, Disease, and Evolutionary Rescue

In 2015, an epidemic of unknown identity eliminated more than half of the population of the critically endangered saiga antelope (*Saiga tatarica*), in the short span of three weeks. While the causative agent was ultimately identified as a species of *Pasteurella*, the mechanism by which a normally asymptomatic non-pathogenic bacterium killed at least 130,000 antelopes is still in question. [47] Literature explaining the die-off, or predicting the consequences for the species, remains comparatively limited; the fate of the species remains uncertain, and it may yet face extinction in the coming years.

Disease is rarely responsible for the extinction of a cosmopolitan species; but for already-threatened species like the saiga, it can be one of the most rapid, unpredictable and unpreventable mechanisms of extinction. Disease has been implicated in a handful of notable wildlife extinctions, like that of the thylacine (*Thylacinus cynocephalus*) and Carolina parakeet (*Conuropsis carolinensis*), and has been the definitive mechanism of extinction for species like the eelgrass limpet (*Lottia alveus*). [48] While most diseases co-evolve with their hosts to an optimal virulence that prevents the species from reaching extinction, diseases that can persist in the environment may be released from such constraints and be more likely to evolve "obligate killer" strategies (like that of anthrax [49]). Fungal pathogens in particular tend to grow rapidly in hosts and spread rapidly between them, which can result in population collapses before optimal virulence levels can be attained. [50]

Two notable fungal diseases have recently demonstrated the destructive potential of environmentally transmitted pathogens. Perhaps the most significant example of disease-driven extinctions is the trail of destruction caused by the chytrid fungus *Batrachochytrium dendrobatidis* (Bd). Bd has been found in at least 516 species of amphibian [51] and has driven decline or extinction in least 200 [52], including at least two thirds of the genus *Atelopus* alone. [53] According to some estimates, current extinction rates that amphibians face (largely but not entirely due to chytrid) are roughly 200 times the background rate; including declining species, that estimate is closer to an even more staggering 25-45,000. [54] White nose syndrome (WNS; *Geomyces destructans*), a similar fungal epizootic, has similarly spread through bat populations in the eastern United States, causing widespread population-level dieoffs since the mid-2000s. While white-nose syndrome has yet to drive any entire species to extinction, significant concern remains regarding its ongoing spread; one study in 2010 using population viability analysis suggested a 99% extinction risk for the little brown bat (*Myotis lucifugus*) in under two decades. [55] Even in a best-case scenario where white-nose mortality was reduced to one twentieth of its rate, substantially reducing extinction risk, bats would still be reduced to one percent of their original population size.

White-nose syndrome has also become a potential case study for evolutionary rescue, one of the most controversial phenomena in extinction research. The premise that rare genes for

resistance or tolerance can bring a disease-ridden population back from the brink of extinction has theoretical support, and potentially indicated from the rapid evolutionary response of certain hosts documented throughout the literature [56]. But WNS constitutes one of the most interesting and controversial examples because, while populations show some sign of recovery from the disease, at the time of writing, no definitive genetic mechanism for resistance has been isolated, a necessary component of demonstrating evolutionary rescue from disease-induced extinction. [56] Consequently, speculation about evolutionary rescue is controversial and so far has been conducted in primarily theoretical settings. In an age-structured matrix population model proposed by Maslo and Fefferman, two scenarios for recovery from WNS are considered. [57] In one, bats' adaptive immunity leads to re-stabilization at much lower levels overall, but a much faster recovery to a stable balance of juveniles (J) and adults (A), with subscript t denoting the number of individuals in these two age classes at time t. In that model, in the absence of white-nose,

$$\begin{pmatrix} J_{t+1} \\ A_{t+1} \end{pmatrix} = \begin{pmatrix} 0.95 & 0.35 \\ 0.95 & 0.87 \end{pmatrix} \begin{pmatrix} J_t \\ A_t \end{pmatrix}$$

In a second model they propose, recovery comes not from adaptive immunity but from innate immunity through a genetic mechanism for resistance. In that scenario a robust type (R) is present in the gene pool with frequency p_t and protects from white nose infection; the remainder of individuals are wild type (WT). In the evolutionary rescue model, all individuals have lower survivorship, but wild type bats fare much worse and reproduce at slightly slower rates (imposing strong selection against WT):

$$\begin{pmatrix} J_{t+1} \\ A_{t+1} \end{pmatrix} = p_t \begin{pmatrix} 0.86 & 0.32 \\ 0.86 & 0.78 \end{pmatrix} \begin{pmatrix} J_t^{\mathrm{R}} \\ A_t^{\mathrm{R}} \end{pmatrix} + (1 - p_t) \begin{pmatrix} 0.52 & 0.27 \\ 0.52 & 0.46 \end{pmatrix} \begin{pmatrix} J_t^{WT} \\ A_t^{WT} \end{pmatrix}$$

In this model, an 11-year stabilization period ultimately leads to population recovery with a positive net growth rate (calculated as the dominant eigenvalue $\lambda = 1.05$), potentially saving populations from extinction. Despite the lack of genetic evidence for evolutionary rescue, Maslo and Fefferman propose that observed similarities between the dynamics they observe and real data on white-nose outbreaks suggests that evolutionary rescue may be happening in real time. Other work since has similarly supported the idea that bat populations may be recovering. Validating these results requires that researchers identify genetic variation between populations associated with differential outcomes, and develop models more directly informed by those mechanisms.

2.4 The Metapopulation Scale

Populations rarely exist in isolation, but are often connected to other populations through dispersal processes, creating a *metapopulation*. Metapopulations are considered to be in a relatively constant state of flux, as local extinctions of species in habitat patches are buffered by re-colonization from local dispersal. In this way, dispersal can be beneficial or detrimental

to metapopulation persistence. Under high dispersal, patches become homogeneous and population dynamics tend to become synchronous. This synchrony is destabilizing, in that periods of low population sizes will be experienced by all patches, increasing the likelihood of stochastic extinction of the entire metapopulation. On the other hand, too little dispersal will result in spatial clustering of a species, as the species will be confined to the set of patches that can be successfully reached and colonized and similarly potentially increasing extinction risk. [58, 59]

The importance of dispersal to patch-level colonization and metapopulation persistence highlights that extinction processes occur at two scales: the local patch-level (i.e., a single population in the network of habitat patches) or at the entire metapopulation level (i.e., either through catastrophic events or cascading local extinctions). Extinctions of single patches can occur as a result of demographic, environmental, or genetic stochasticity (addressed in more detail in §2.3), or through extrinsic events related to habitat loss or natural enemies [60]. Metapopulation level extinction can also result from environmental stochasticity at the regional scale [61], provided this stochasticity is spatially autocorrelated, such that it is expected to promote synchronous dynamics among habitat patches [62].

Basic Metapopulation Models and Extinction

In the classic metapopulation model described by Richard Levins, the balance between patch colonization (c) and local extinction (e) determines patch occupancy dynamics. In this case, local habitat patches are either occupied or unoccupied, and both patch number and the spatial orientation of patches are undescribed. Dispersal among habitat patches can rescue patches from extinction, or allow for the recolonization of extinct patches. All patches are treated as equal, so that any patch is suitable for a species, and (as a simplifying assumption) all habitat patches can be reached from all other patches. This simplified representation treats space as implicit, and patch quality and size as constant; rather than an explicit population size, patch occupancy is just a 0 or 1 state. The dynamics of the proportion occupied patches, P, are given by a differential equation:

$$\frac{dP}{dt} = cP(1-P) - eP \tag{2.14}$$

In that equation, extinction is a random process for every occupied patch that is entirely independent of the state of the system. In contrast, colonization rates depend both on the fraction of occupied and unoccupied patches, as emigrants move from occupied patches to re-colonize unoccupied ones. The balance between the two processes of extinction and colonization determines long-term persistence of the metapopulation; [63] that is, a necessary condition for metapopulation persistence in this model is

$$\frac{e}{c} < 1$$

At a non-trivial equilibrium, the patch occupancy is given as

$$\hat{P} = 1 - \frac{e}{c} \tag{2.15}$$

This suggests that the equilibrium fraction of occupied patches is a simple function of colonization (c) and extinction (e). If extinction rates are greater than zero, this implies that the equilibrium occupancy is less than one even if colonization exceeds extinction; that is to say, not every patch will ever be stably filled if extinction is nontrivial. This shows that even a metapopulation in equilibrium is still in a constant state of patch-level flux. In real applications, this implies that just because a patch of habitat is empty, that may not imply it is uninhabitable; and similarly, just because a population goes extinct, it may not be indicative of broader declines or instability.

While admittedly a simple representation of a metapopulation, the Levins model can yield important insights into spatial population dynamics [64]. For instance, the mean time to extinction of any given population/patch is the inverse of the rate (i.e., $T_E = 1/e$), providing a link to the models at the population scale discussed above. We can take the Levins model a step further to explicate the relationship between patch occupancy and overall mean time to extinction T_M at the metapopulation scale. Starting with the assumption that the total H patches each have their own average extinction time T_L (which should be $1/e$),

$$T_M = T_L \exp\left(\frac{(\hat{P}H)^2}{2H\left(1-\hat{P}\right)}\right)$$

Consequently, using Eq. 2.15, we can also express T_M as

$$T_M = T_L \exp\left(\frac{H}{2}\left(cT_L + \frac{1}{cT_L} - 2\right)\right)$$

showing that metapopulation extinction time increases exponentially, not linearly, with the MTE of individual habitat patches. [65]

The simplicity of the Levins model has resulted in a sizable body of literature surrounding and extending the model. For instance, in the original Levins' model all patches are equidistant from one another, identical in quality, and can only be in one of two potential states (occupied or unoccupied), but each of these conditions is frequently adjusted in derivative stochastic patch occupancy models (SPOMs). Researchers have shown that despite the simplicity, Levins-type dynamics can emerge from more complicated stochastic metapopulation models [64], and extensions of the Levins model continue to provide insight into the influence of habitat patch size and topography (i.e., spatial orientation of habitat patches) on metapopulation persistence [66].

Island Biogeography and Metapopulation Capacity

A simple extension of the Levins model considers a set of spatially explicit patches of variable size, where a distance matrix D describes the distance between all patches in the metapopula-

tion. The model borrows elements of Macarthur and Wilson's *Theory of Island Biogeography* [67], such that distance between patches (D_{ij}) and patch area (A_i) influence extinction and colonization processes, where the patch extinction rate scales with patch area ($e_i = e/A_i$), and colonization (c_i) becomes a property of distance (D_{ij}), patch area (A_i), and dispersal rate (α) where

$$c_i = \sum_{j \neq i} e^{-\alpha D_{ij}} A_j p_j(t)$$

This suggests that the mean time to extinction of a habitat patch ($1/e_i$) is determined by the area of the patch. This makes the occupancy probability of each patch in the metapopulation, described in terms of matrix M

$$M_{ij} = e^{-\alpha D_{ij}} A_i A_j$$

and the leading eigenvalue of this matrix M describes the persistence of the metapopulation (also know as *metapopulation capacity*[68] or λ_M). The condition for metapopulation persistence is that the dominant eigenvalue of M must be greater than the ratio between extinction and colonization rates:

$$\lambda_M > e/c$$

While spatially explicit, this approach assumes that dispersal among habitat patches is determined by patch area and distance to other patches, ignoring population dynamics in each patch. However, since habitat patches vary in their size and connectedness to other patches, it is possible to determine the relative importance of each habitat patch to metapopulation persistence in this framework [68, 69], potentially informing conservation and management decisions [70].

Incorporating Patch Dynamics

The above extension of the Levins' model allows for patches to vary in size and connectedness. Another extension is to consider the abundances of habitat patches within the metapopulation, thus considering the dynamics of each patch, and the effects of dispersal among local populations [71].

$$N_i(t+1) = R_i(t) N_i(t) e^{-N_i/K}$$

This expression assumes that the growth rate of each habitat patch is R_i, and that the carrying capacity is a constant K. If we assume that the population growth rates (r_i) are independent and identically-distributed Gaussian random variables, this causes R_i values to be log-normally distributed, and allows us to define persistence thresholds for the metapopulation based on the variance in the population growth rates r_i. The threshold for metapopulation persistence relies on exceeding a threshold value ($\sigma_{threshold}$) in terms of the

variance among local patch population growth rates (r_i). If μ is the mean local population growth rate over time, this threshold is

$$\sigma_{threshold} > \sqrt{2|\mu|}$$

This model can be extended to yield many interesting conclusions. For instance, if populations have influence on where their offspring go, population growth rates may be maximized by seeding offspring in less than suitable "sink" habitat if habitat quality fluctuates with time, and when the "source" habitat occasionally experiences catastrophes [72]. The complexity of metapopulation dynamics in the face of environmental stochasticity, variable patch quality, dispersal, and competition has fueled expansive theoretical work [73, 74]. An obvious next step is to scale from single species metapopulations to multi-species communities (i.e., metacommunities), which allows for the modeling of how species interactions, predator-prey dynamics, and community assembly relate to persistence [75].

2.5 The Species Scale

Extinction is defined at the scale of the species, but it is also at this level of taxonomic resolution that it is perhaps hardest to quantify—and, to summarize—due to considerable diversity of approaches and applications. We explore in this chapter two applied extensions of that body of theory, corresponding to two common quantitative frameworks for species-level extinctions. In the first, the complete loss of suitable habitat leads to an inevitable—if not immediate—extinction. Species can escape extinction through three primary channels: acclimation, adaptation, and migration. Species distribution models are often used to calculate extinction risk at the community scale in that framework (described in greater detail below), but they can only at best include the last of those three rescue processes. Evolutionary models, on the other hand, can link demography and genetics to the overall risk of extinction in a changing environment; we explore that application here in the context of both adaptation and phenotypic plasticity.

The second framework is based in the notion that population extinctions become species extinctions; and so the framework for population (and metapopulation) viability analysis described above acts as a sufficient method for estimating species extinction risk. In many cases, that may be a safe assumption, as near-extinction species are reduced down to a single persistent population or a handful in isolated refugia. But in real applications, persistence in small isolated refugia may be difficult to study, or even observe with any regularity; consequently, an entire body of literature has been developed to relate extinction risk to the sightings of rare species. That body of theory allows two applications: the posthumous assignment of extinction dates to extinct species, and sighting-based hypothesis testing for a species of unknown extinction status. We explore both applications below.

Adaptation and Plasticity in a Changing Environment

Bounding uncertainty is the seminal challenge in extinction research, and in the real world, species' potential to acclimate and adapt to changing environments confers an unknown degree of robustness that may give species a chance at evading extinction. As discussed above, evolutionary rescue has been a particularly tantalizing—and controversial—idea in the context of disease research. But more broadly, evidence suggests that extinction risk is heavily complicated by species' variable ability to track changing environments.

Most models of evolutionary rescue approach the problem by explicitly modeling fitness curves and the speed of natural selection. In a foundational paper by Gomulkiewicz & Holt [76], population size N_t changes over time in response to its mean fitness W_t such that

$$N_t = W_{t-1}N_{t-1} = \prod_{i=1}^{t-1} W_i N_0$$

If fitness is below one (i.e., populations are reproducing at a rate below replacement), then the population will tend towards extinction. The model they present uses a pseudoextinction threshold N_c such that if the initial fitness W_0 is held constant over the entire interval,

$$T_E = \frac{\ln N_c - \ln N_0}{\ln W_0}$$

Without adaptation (i.e. W_t increase above $W_0 < 1$), the population declines to extinction. To model adaptation, Gomulkiewicz & Holt assume that environmental change begins at time 0, adapting a system of equations for describing natural selection on a single phenotypic trait originally proposed by Russell Lande. [77] In that notation, the trait z has an optimum phenotype z_{opt}. The population mean phenotype is expressed as d_t, the distance of the average z from z_{opt} at each timestep, with an initial value d_0. As for any quantitative trait, individual phenotypic values z are normally distributed around the population mean with some variance σ_z^2:

$$z_t \sim \mathcal{N}(d_t, {\sigma_z}^2)$$

The corresponding fitness function with width ω_z is expressed as a bell curve around the optimum:

$$W(z) = W_{\max} e^{-z^2/2\omega_z}$$

where $W_{\max}$ is the fitness of z_{opt}. The width of the fitness function (which can be interpreted as the strength of selection), the existing variance in the trait, and the distance from the optimum, determine how quickly the population evolves; the changing fitness of the population can be expressed as:

$$W_t = W_{\max}\sqrt{\omega_z/(\sigma_z^2 + \omega_z)}\; e^{-d_t^2/({2\sigma_z}^2 + 2\omega_z)} \tag{2.16}$$

Even a population with a mean at z_{opt} does not have perfectly maximized fitness, because of the variance around the mean; the actual growth rate of the population when $d_t = 0$ can be expressed as

$$\hat{W} = W_{\max}\sqrt{\omega_z/(\sigma_z^2 + \omega_z)}$$

This provides a clear way to simplify Eq. 2.16:

$$W_t = \hat{W}e^{-d_t^2/(2{\sigma_z}^2+2\omega_z)}$$

In this expression, the changing fitness of the population is expressed only as a function of the optimum and the strength of selection on the trait z.

How does the actual distribution of phenotypes change over time? In real systems, evolution is seldom a direct progression towards the optimum, even under hard selection with ample genetic variation. If the trait z has a heritability h^2 (where a heritability of 1 means the trait is perfectly heritable, and 0 would indicate perfect plasticity or no genetic basis), Gomulkiewicz & Holt define a scaleless "evolutionary inertia"

$$k = \frac{\omega_z + (1-h^2){\sigma_z}^2}{\omega_z + {\sigma_z}^2}; 0 \leq k \leq 1$$

which in turn simplifies how fast the population shifts towards its optimum phenotype:

$$d_t = k^t d_0$$

Together, this set of equations produces the governing expression for the system:

$$t\ln\hat{W} - \frac{{d_0}^2}{2(\omega_z + {\sigma_z}^2)}\frac{1-k^{2t}}{1-k^2} = \ln\frac{N_c}{N_0}$$

If this equation has no roots when solving for t, then this indicates the population will fall and rise without any real extinction risk; but when it does, the roots are estimates of the time until the population falls below the critical threshold (T_E) and the time until recovery could be evolutionarily possible (T_P in their notation, where N_t passes back above N_c). The interval between these two values is characterized by a small population that, due to demographic stochasticity, would require much more intensive conservation efforts (e.g., managed *ex situ* breeding) than normal to possibly survive that interval. The time to recovery (growth switches from negative to positive even though $N_t < N_c$) is

$$T_R = \frac{1}{\ln k^2}\left(\ln\ln\hat{W} - \ln\frac{{d_0}^2}{2(\omega_z + {\sigma_z}^2)}\right)$$

From this expression, Gomulkiewicz and Holt derive a useful finding: "t_R increases logarithmically with the degree of initial maladaptation ... but is independent of the initial population density." Or, to rephrase: the possibility and speed of evolutionary rescue depends on

the initial phenotypic distribution, the evolutionary inertia, and the speed of selection, but is scale invariant across population sizes; even small populations with high enough genetic diversity and low inertia can be rescued by evolutionary rescue.

The model developed by Gomulkiewicz and Holt sets useful theoretical bounds on the genetically-coded evolution of a trait; but in the real world, phenotypic plasticity complicates this pattern, and presents one of the hardest challenges for predicting how species might escape extinction. In a similar model developed by Chevin *et al.* [78], the trait in question z has a developmental trajectory with both a genetic component and the potential for phenotypic plasticity in response to an environmental variable ϵ. Their model uses a "reaction norm" approach to plasticity (popularized by Schlichting, Pigliucci and others [79]), breaking down that phenotypic trait into an adaptive genetic component a and a plastic component b that responds to the environmental gradient. They express the distribution of the phenotype $p(z)$ at generation n in an environment changing at rate $\epsilon(t) = \eta t$ as:

$$p(z) \sim \mathcal{N}(\bar{z}, {\sigma_z}^2)$$

Here the population mean $\bar{z}$, expressed in terms of the generation time T under the assumptions that i) developmental plasticity takes effect at time τ during ontogeny and ii) the strength of plasticity b (the slope of a phenotypic reaction norm), takes the form

$$\bar{z} = \bar{a} + b\eta T(n - \tau)$$

Assumption (ii) is of course a limiting one, given that plastic reaction norms are in fact evolvable; but extensions of quantitative theory that incorporate this idea are underdeveloped. We also assume that the variance associated with z has both environmental and genetic components: i.e.,

$$\sigma_z^2 = {\sigma_a}^2 + {\sigma_e}^2$$

Assuming there is an optimum phenotype $\theta = B\epsilon$, where B is the optimal rate of change to plastically track the changing environment, Gomulkiewicz and Holt define a changing population size with a maximum growth rate $W_{\max}$, such that

$$W(z) = W_{\max} \exp\left(-\frac{(z-\theta)^2}{2\omega_z} - \frac{b^2}{2\omega_b}\right)$$

where both ω's represent the strength of stabilizing selection (the width of fitness curves, comparable to above). From there they make the link to overall population dynamics, where the intrinsic rate of growth of the population r can be scaled with generation time and related to selection on z as

$$r = \frac{\ln(W)}{T} = \frac{\ln(W_{\max})}{T} - \frac{\ln(1 + \sigma_z^2/\omega_z) + b^2/\omega_b}{2T} - \frac{(\bar{z} - \theta)^2}{2T}\gamma$$

where γ, the strength of stabilizing selection, is given by

$$\gamma = \frac{1}{\omega_z + \sigma_z^2}$$

The first two terms become the maximum possible growth rate r_{max} if z reaches the optimum θ:

$$r_{max} = \frac{\ln(W_{\max})}{T} - \frac{\ln(1 + \sigma_z^2/\omega_z) + b^2/\omega_b}{2T}$$

From this expression for population dynamics, Chevin *et al.* derive a formula for the critical rate of environmental change, above which ($\eta > \eta_c$), plasticity and adaptation combined still fail to prevent extinction (recalling that B is the optimal rate of change to plastically track the changing environment and b is the slope of the phenotypic reaction norm):

$$\eta_c = \sqrt{\frac{2r_{\max}\gamma}{T}}\ \frac{h^2\sigma_z^2}{|B - b|}$$

From this expression, it is very easy to determine the long term tendency of the population to extinction or survival as a function only of the degree of plasticity and the associated strength of costs (ω_b). The greater the extent of plasticity, the more the costs of plasticity separate out population trajectories; but when plasticity has a weak slope, the extinction isoclines converge towards the same threshold. This conceptualization of adaptation to environmental change as a single-trait system with readily measured costs of adaptive plasticity is obviously an idealization, but also clearly illustrates a number of important points. While adaptive genetic variation has a clear direct relationship to evolutionary rescue, plasticity also plays an important role; and quantifying plasticity without quantifying its costs can provide a misleading perspective on the feasibility of adaptation and acclimation.

Is Evolutionary Rescue Real?

Evolutionary rescue is not a "silver bullet," and the application of evolutionary theory to real populations and metapopulations is far from straightforward. For one thing, evolutionary rescue requires a sufficiently large population that a species is buffered against demographic and environmental stochasticity long enough for higher-fitness phenotypes to become predominant. [80] Additional complications include, but are not limited to:

- **Initial environmental conditions.** Bell and Gonzalez showed that populations that begin at intermediate stress levels may react the slowest to environmental "deterioration," producing a U-shaped curve in adaptive rescue. [81] They explain this as a product of two competing processes driving evolutionary rescue: as baseline stress increases, overall mutation rates decline, but the proportion of beneficial mutations (or, perhaps more accurately, the associated fitness differential) increases. Populations beginning in "mildly stressful conditions" may simply be at the low point of both processes. Bell and Gonzalez similarly show that populations with a history of minor

environmental deterioration have a much greater probability of evolutionary rescue in a fast-changing environment.

- **The velocity of environmental change.** As Chevin *et al.*'s model highlights, environmental changes that are too rapid almost invariably drive species to extinction, when selection simply cannot operate fast enough to keep pace; this finding is readily confirmed in environmental settings. Rapid environmental changes can also functionally reduce mutation rates at a population scale. A study of *E. coli* by Lindsey *et al.* showed that "The evolutionary trajectory of a population evolving under conditions of strong selection and weak mutation can be envisioned as a series of steps between genotypes differing by a single mutation," and some "priming mutations" may be necessary to arrive at further genotypic combinations with substantially higher fitness. [82] Consequently, if environmental changes are too rapid, higher fitness genotypes may be "evolutionary inaccessible."

- **Dispersal rates and metapopulation connectivity.** Simulated metapopulation models by Schiffers *et al.* showed that higher dispersal rates can severely limit the propensity of populations to experience local adaptation, especially in a heterogeneous environment (a phenomenon they refer to as "genetic swamping"), and thereby potentially limit evolutionary rescue. [83] However, for an entire species to persist, intermediate (local) dispersal may be necessary to allow adaptive mutations to spread, a finding shown experimentally by Bell and Gonzalez.

- **Linkage disequilibrium.** Schiffers *et al.*'s study, which simulated genomes in an "allelic simulation model," produced an unusual result suggesting that linkage between adaptive loci may not actually increase the rate of adaptation. The interaction this could have with the "priming mutation" process is complex and poorly explored in a theoretical context.

A final important consideration should be made with regard to what Schiffers *et al.* distinguish as *complete* vs. *partial evolutionary rescue.* In their models, they find that when adaptive traits originated but spread poorly (as a combination of linkage disequilibrium, habitat heterogeneity, and dispersal limitations), it substantially reduced population sizes and ultimately produced an "effective reduction in the suitable habitat niche." This type of partial evolutionary rescue could be most common in real-world scenarios, where adaptation in larger populations experiencing the slowest rates of environmental change may allow persistence, but not maintain a species throughout its entire range, and may still be followed by a substantial reduction in overall habitat occupancy.

If current research on global climate change is any indication, this type of partial evolutionary rescue may ultimately be a poor buffer against extinction. Climate change may set the events of an extinction in motion, but research suggests that habitat loss from climate change is rarely the direct and solitary causal mechanism of an extinction. [84] Instead, climate change may reduce a population to small enough levels at which other mechanisms

drive extinction. Small populations are especially susceptible to stochastic crashes in population size, and may also be especially susceptible to stochastic collapse due to other factors within-species (Allee effects in breeding, inbreeding) or from interactions with other species (competition, invasion, disease). Ultimately, the synergy between these drivers may produce a greater overall extinction risk that many modeling approaches might not directly quantify, but that could be most likely to drive species to extinction, and drive ecosystems into novel assemblages. [85]

After Extinction: Lazarus Species, Romeo Errors, and the Rarest Birds in the World

The job of conservation biologists and extinction researchers is far from over after the extinction of a species. The *autoecology* of an extinct species (its basic biology, ecology, natural history, distribution and other species-level characteristics) often becomes a permanent unknown, assumed to be lost to the annals of history. But as statistical tools for ecological reconstruction become more sophisticated, researchers have the chance to explore basic questions about extinction in retrospect. In particular, the same body of theory that governs the timing of extinction in a declining population can be applied in a retrospective sense as well, to estimate the likely extinction date of a species. (Or, more formally, the estimation of the MTE from a given point can be used to pinpoint T_E , even with the same data, after extinction has already occurred.) These methods have been used both for ancient species like the megalodon [86], and for more recent extinctions like that of the dodo [87]. But perhaps most interestingly that theory can be applied when the uncertainty bounds on T_E contain the present date, meaning that the extinction of a species is not taken as a certain part of history. Even ancient "Lazarus species" can be rediscovered, like the coelacanth, believed to have gone extinct 66 million years ago but rediscovered in the last century. How can we confidently say the coelacanth continues to exist, but the megalodon is likely to never be rediscovered?

Statistical Methods for the Sighting Record

Once a species is suspected to be extinct, at what point do we stop looking for them? With limited resources for conservation, trying to find and conserve a species that is no longer around wastes resources better used elsewhere; but making a Type I error and assuming a species is falsely extinct (and abandoning conservation efforts) can lead to a "Romeo Error," whereby giving up on the species can lead to actual extinction. [88] Since 1889, 351 species thought to be extinct have been "rediscovered,"[89] highlighting just how big of a problem this may be. In order to answer these questions, determining the probability that a species is still extant, despite a lack of recent sightings, is an important step in making evidence-based decisions conservation managers must make about allocating resources.

But how do we determine the likelihood that a species is extinct? How long does it have to be since the last time an individual was seen before we can say, with some certainty,

that the species is gone? The most obvious step is to assemble all available evidence from when the species was around. The first place to look is in the specimen record, which conventionally acts as the "gold-standard" of evidence. However, other data can be brought to bear, including observations, photos, and audio recordings. All these forms of evidence are collectively referred to as *sightings*. For a dataset of sightings $\mathbf{t} = (t_1, ..., t_n)$, perhaps the simplest approach is to wait at least as long as the last interval during which the species was apparently absent before declaring a species extinct. One could formalize the estimate of the extinction date, T_E, as:

$$\widehat{T_E} = t_n + (t_n - t_{n-1})$$

This approach, formalized by Robson and Whitlock[90], is accompanied by a $(1 - \alpha)\%$ confidence interval with a lower bound at the last sighting t_n and the upper bound

$$T_E^u = t_n + \frac{1-\alpha}{\alpha}(t_n - t_{n-1})$$

and accompanying p-value for testing the hypothesis that the species is extinct at the current time T:

$$p = \frac{t_n - t_{n-1}}{T - t_{n-1}}$$

The reasoning behind this method is fairly sound: if a large gap exists between the last two sightings, conservation biologists should wait at least that long before pronouncing a species certain to be extinct. But this estimator is also severely conservative, and has very limited theoretical grounding.

In 1993, Andrew Solow developed a more explicitly probabilistic approach [91], which assumes sightings are generated by a random process with by a fixed sighting rate m that becomes 0 at T_E, the true date of extinction. The probability of the data conditional on a current time T and an extinction date T_E, is

$$P(T_n \leq t_n | T_E \geq T) = (t_n/T)^n$$

In that light, Solow says, hypothesis testing is easy: against the null hypothesis that extinction has yet to happen (i.e., $T_E > T$), we can test the alternate hypothesis that the species is extinct ($T_E < T$). For a given last sighting at T_N, we can provide a p-value for the test with desired significance level α equivalent to

$$P(T_N \leq \alpha^{1/n} T | T_E < T) = \alpha (T/T_E)^n$$

for values of $\alpha^{1/n} T < T_E < T$; for values of T_E lower than or equal to that critical value $\alpha^{1/n} T$, the value of that P is equal to 1 and the null hypothesis is rejected with full certainty. Solow explains, by way of example, that with 10 sightings and 95% confidence, the critical value of T_E/T is 0.74, and so the null hypothesis is sure to be rejected (extinction is confidently confirmed) if the true extinction date occurs within the first 74% of the $(0, T)$ window.

Based on this approach, the maximum likelihood estimate $\widehat{T_E}$ would be t_n, but this is clearly biased, and performs poorly as an estimation method. Instead, he suggests an alternate non-parametric estimator [92]:

$$\widehat{T_E} = \frac{n+1}{n} t_n$$

And, in addition, a $1-\alpha$ upper confidence interval bound:

$$T_E^u = t_n/\alpha^{1/n}$$

Solow also proposed a foundational Bayesian approach based on the likelihood a given sighting rate m would generate an observed density of data. Hypothesis testing in the Bayesian format where the likelihood of the sighting data given H_0 (the species is extant at time T) is

$$\int_0^\infty m^n e^{-mT} dP(m)$$

and given H_A (the species went extinct at T_E) is

$$\int_0^\infty m^n e^{-mT_E} dP(m)$$

From those and other assumptions, he derived the Bayes factor for the hypothesis test:

$$B(\mathbf{t}) = \frac{n-1}{(T/t_n)^{n-1} - 1}$$

In Bayesian statistics, the Bayes factor is used to express the relative support between these two hypotheses. It bypasses the problem of setting a prior on the data or, in fact, of the two hypotheses; and instead, just expresses the posterior:prior odds of H_0. If H_0 is small ($B << 1$), that suggests there is strong evidence against the null hypothesis.

Do these approaches make sense? If an extinction happens abruptly on the scale of sightings data (say, an epidemic wipes a species out within a year), then sighting rates might remain relatively constant throughout the sighting record. Similarly, applying this method to paleontological records may make sense, as prior information about variation in specimen preservation might be limited (and so a constant rate parameter is the best possible prior). But there are also a number of situations where the constant sighting rate m simply does not suffice. Lessons from population ecology remind us that extinction is, at its most fundamental scale, a process of declining abundance. If sightings are proportional to abundance (which they generally are), replacing m with a non-constant function has the potential to sharply refine the process of extinction date estimation.

Two additional methods have been suggested by Solow to account for the changing rate of sightings. Both assume that sightings are a declining process, which will make at least some of the above estimators prone to Type I errors. The first method assumes sighting

rates decline exponentially at a rate β, so that the sighting density for $0 \leq t \leq T_E$ can be expressed as:

$$f(t) = \frac{\beta e^{-\beta t}}{1 - e^{-\beta T_E}},$$

In this model, if we express

$$s = \sum_{i=1}^{n} t_i$$

and k is the integer part of s/t_n (which we can write as $[s/t_n]$), then the estimated extinction date can be given as:

$$\widehat{T_E} = t_n + \frac{\sum_{i=0}^{k} (-1)^i \binom{n}{i} (s - it_n)^{n-1}}{n(n-1) \sum_{i=0}^{k-1} (-1)^i \binom{n-1}{i} (s - (i+1)t_n)^{n-2}}$$

The p-value is given as $p = F(t_n)/F(T)$ where

$$F(x) = 1 - \sum_{i=1}^{[s/x]} (-1)^{i-1} \binom{n}{i} \left(1 - \frac{ix}{s}\right)^{n-1}$$

The confidence interval can be determined using this p-value. However, computationally, that upper bound does not always exist–a major problem with this method.

In contrast, the second and far more complex method, implemented by Roberts and Solow in their 2003 study of the dodo [87], accounts for the fact that the last few sightings of the species should, in most circumstances, follow a Weibull distribution. The method, *optimal linear estimation* (OLE), estimates T_E through linear algebra, with

$$T_E = \sum_{i=1}^{k} w_i t_{n-i+1}$$

$$w = (e' \Lambda^{-1} e)^{-1} \Lambda^{-1} e$$

$$e = \begin{pmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} \qquad \text{(dimension } k\text{)}$$

and Λ is a k by k matrix, for which

$$\Lambda_{ij} = \frac{\Gamma(2\hat{v} + i)\Gamma(\hat{v} + j)}{\Gamma(\hat{v} + i)\Gamma(j)}$$

$$\hat{v} = \frac{1}{k-1} \sum_{i=1}^{k-2} \ln \frac{t_n - t_{n-k+1}}{t_n - t_{i+1}}$$

The OLE's upper 95% confidence interval is given by:

$$\widehat{T_{ci}^{u}} = T_n + \frac{T_n - T_{n-k+1}}{c(\alpha) - 1}$$

$$c(\alpha) = \left(\frac{-log(\alpha/2)}{k}\right)^{-\hat{v}}$$

The OLE method has been recorded as one of the most successful methods available for predicting extinction [93], and has the added bonus of being adjustable through sensitivity analysis to examine how different extent of sighting data changes the overall estimate.

Case Study: Spix's Macaw

Perhaps the most fruitful body of research concerning extinction date estimation has been within ornithology, where data on the last sightings of rare species is often more available than for other groups, due to tremendous global interest in bird sightings and observation by non-scientists. The most popular methods for sighting date research have often been developed in association with data on notable extinct birds, including the dodo[87] and the ivory-billed woodpecker[94]. In fact, one of the most expansive reviews of sighting date estimators, conducted by Elphick and colleagues, estimated the extinction date of 38 extinct or near-extinct birds from North America (including Hawaii, a hotspot of bird extinction). [95] But for rarer birds around the world, basic data on their extinction may be somewhat more lacking.

One such bird, the Spix's macaw (*Cyanopsitta spixii*) has been called "the world's rarest bird" [96] and has been the subject of two popular animated movies (Rio and Rio 2). Currently, Spix's macaw is considered critically endangered (possibly extinct in the wild) by the IUCN, with a small number of captive individuals (∼130) found around the world. Not seen in the wild since 2000, a video of a Spix's macaw in Brazil made headlines in 2016. The video was subsequently examined by ornithologists and the consensus that the bird was, in fact, a Spix's macaw, though many believe it was likely an escaped captive bird.

Sightings of the Spix's macaw are sporadic, and after the first known specimen being shot in 1819 by Johann Baptist Ritter von Spix (though he believed the bird to be a Hyacinth Macaw), it was not recorded again until a wild-caught individual was procured by the Zoological Society of London in 1878. Collecting sighting records of the Spix's macaw relies mostly on data from trappers/poachers and inferring data from captive individuals. Given the illicit nature of wildlife poaching, better data may exist in the husbandry records of the wild-caught individuals currently in captivity, but those data are not freely available. Verifiable observations are few and far between, as this species was not subject to any intensive study or searches until the mid-1980s, when only a handful of individuals were found and, of those remaining, most were caught by poachers.

For this case study, we collected sighting and specimen data from GBIF (Global Biodiversity Information Facility; www.gbif.org) and Juniper's authoritative book on Spix's macaw.

We found physical evidence (specimens / wild-caught captive birds) for sightings in the years: 1819, 1878, 1884, 1901, 1928, 1954, 1977, 1984, 1985, 1986, and 1987. Due to their rarity and the demand for them, we assumed individuals were caught in the wild the same year they were procured by the receiving institution / zoo. We considered all observations of the Spix's macaw reported in Juniper's book to be as uncontroversial as physical specimens, as there aren't many and these few have been rigorously scrutinized: 1903, 1927, 1974, 1989, 1990, and 2000. Our only controversial sighting is the recent video taken in 2016, which we omit from the model. By eliminating the controversial sighting (in analyses 1 and 2), we inherently test a methodological question: would extinction date estimators have pronounced the apparently-extant species dead?

Our analysis was conducted using the beta version of `R` package `sExtinct`, which allows a handful of different extinction analyses to be implemented (and we encourage prospective users to test the demos available with the package). [97] Our analysis uses two of the most common methods. First, we used the original Solow maximum likelihood approach, plotting the probability of persistence in Figure 1.2. The maximum likelihood estimates are given in that method as:

- Specimens only (1819-1987): $T_E = 2040$
- Uncontroversial sightings (1819-2000): $T_E = 2035$
- All sightings (1819-2016): $T_E = 2052$

That method suggests, even with the most limited dataset, that the species still appears to exist. In contrast, the OLE method tells a different story:

- Specimens only: $T_E = 1988$ (95% CI: 1987 - 2006)
- Uncontroversial sightings: $T_E = 2002$ (95% CI: 2000 - 2018)
- All sightings: $T_E = 2021$ (95% CI: 2016 - 2045)

All things considered, both analyses suggest at least a marginal chance the 2016 sighting may have been legitimate, and there is a possibility that a wild population of Spix's macaws may be out there, yet undiscovered in the Amazon rainforest. But, the OLE method—for all its documented strength as an approach—would likely have been far hastier to dismiss the species as extinct before its 2016 "rediscovery." Furthermore, with some researchers hoping to use extinction date estimators as a method of Red Listing, we note that the OLE currently only predicts another five years of persistence for the species, and the rarest bird in the world clearly remains a severely threatened species.

Hope Springs Eternal: Addressing False Sightings

Consider the plight of the ivory-billed woodpecker (*Campephilus principalis*), a charismatic and iconic part of the North American fauna. The ivory-billed woodpecker's decline was gradual, and unlike its gregarious and easily-spotted compatriots (such as the passenger pigeon, *Ectopistes migratorius*, or the Carolina parakeet, *Conuropsis carolinensis*, both extinct in a similar time period), sightings of the woodpecker were already rare previous to its decline. So while the bird's last "credible" sighting was in 1944, the precise date of its extinction remains controversial, and some believe the bird still exists based on unverified observations as recent as 2004 (with audiovisual evidence reviewed in a highly controversial 2005 paper in *Science* [98]). These controversial observations led to one of the most costly surveys in history, yet yielded no new evidence. In some circles, the search continues; in 2016, two ornithologists—Martjan Lammertink and Tim Gallagher—traveled through Cuba searching for remaining populations of the elusive woodpecker. Was Lammertink and Gallagher's search justified from a statistical standpoint? And perhaps, more importantly, how can we address the problem of inaccurate sightings?

Not all sightings are created equal. Holding a dead body of an individual of the species constitutes good evidence the species was present the year the specimen was collected; but if some person claims they saw an extremely rare species with no corroborating evidence, they may have misidentified the individual, or in some cases, even lied, meaning that this sighting could be be invalid. Roberts *et al.* found that extinction date estimators are sensitive to the data used, and can, unsurprisingly, lead to very different estimates of extinction dates. [99] They partitioned sighting data into three categories: 1) physical evidence, 2) independent expert opinion, and 3) controversial sightings in order of certainty. They found that adding independently-verified observations to the analysis can sometimes lead to earlier predicted extinction times, since the "gaps" within the sighting record are closed up, whereas, by nature, later controversial sightings, if treated as legitimate (i.e., on par with physical evidence), can greatly push the estimates of extinction to later years.

To account for this uncertainty, a few approaches have been proposed recently. These approaches largely expand on Solow's 1993 Bayesian equation above, modified to consider multiple levels of uncertainty in the data. [94, 100, 101] In the most advanced of these models, valid and invalid sightings are generated by separate Poisson processes. If valid sightings occur at rate Λ and invalid sightings at Θ, the proportion of valid sightings is

$$\Omega = \frac{\Lambda}{\Lambda + \Theta}$$

The sightings data can be split into certain data (t_c, with n_c certain records) and uncertain data (t_u, with n_u uncertain sightings, and $n_u(\hat{T}_E)$ sightings before $\hat{T}_E$). The conditional likelihood of the data is that of the two datasets multiplied:

$$p(t|\hat{T}_E) = p(t_c|\hat{T}_E)p(t_u|\hat{T}_E)$$

$$p(t_c|\hat{T}_E) = \frac{(n_c-1)!}{(\hat{T}_E)^{n_c}}$$

$$p(t_u|\hat{T}_E) = \int_0^1 \omega^{-n_u}(1-\omega)^{n_u-n_u(\hat{T}_E)}\left(\hat{T}_E + \frac{1-\omega}{\omega}T\right)^{-n_u} d\omega$$

where ω is a dummy variable representing the certainty rate Ω. Bayes' theorem gives the probability the species went extinct in the observation interval $(0, T]$, an event E with probability

$$p(E|t) = \frac{p(t|E)p(E)}{p(t)} = \frac{p(t|E)p(E)}{p(t|E)p(E) + p(t|\bar{E})(1-p(E))}$$

The conditional likelihood of the data is

$$p(t|E) = p(t|\hat{T}_E)p(\hat{T}_E)$$

and conversely $p(t|\bar{E})$ is evaluated using the same function but replacing $\hat{T}_E$ with T. The prior probability of the extinction date $p(\hat{T}_E)$ is a key part of successfully implementing Bayesian analyses, and has a significant effect on the estimated extinction date [102]. Solow & Beet (2014) suggest three possible priors: uniform, linear, or exponential decline after the last certain sighting.

As previously noted, setting a prior probability for $p(E)$ is even more challenging. Instead of explicitly calculating the probability the species is still in existence, one can simply calculate the Bayes factor

$$B(t) = \frac{p(t|E)}{p(t|\bar{E})}$$

This bypasses the problem of assigning a prior probability of extinction $p(E)$. However, instead, it can be set uninformatively as 0.5 (equal probability extant or extinct), allowing explicit calculation of $p(E|t)$.

Solow & Beet's model is one of a handful of models that all use Bayesian approaches to estimate the extinction date, and test whether a species is extant. For an overview of the assumptions and relative strengths of these approaches, see Boakes *et al.*[103] We note that while some more complex methods exist that, for instance, assign different intermediate levels of certainty to different kinds of evidence [100], this may ultimately be superfluous. In many cases, the expenditure and effort required to obtain expert opinions may only have a marginal benefit, contributing little extra certainty to the models [104]. Consequently, the choice of model should depend on the available data, the operational power of any given study, and the degree of certainty needed for decision making.

The Ivory-Billed Woodpecker, and the Hunt for More Lazarus Species

To briefly reconsider Lammertink and Gallagher's continuing search for the ivory-billed woodpecker: regardless of how the sighting record for the ivory-billed woodpecker is analyzed, all indications point to an extremely low likelihood that the species is extant [95, 105, 94]. In the work of Elphick *et al.*, estimates based on physical evidence suggested a T_E of 1941 (upper 95% CI: 1945) and including expert opinion sightings only moves T_E towards 1945 (upper 95% CI: 1948). Solow & Beet reanalyzed this problem with their Bayesian models that differentiate between certain and uncertain sightings. With a uniform prior on T_E over 1897-2010, if valid and invalid sightings are treated as generated from the same Poisson process, the calculated Bayes factor of 0.13 strongly suggests the species persists; but in the model that treats the processes as separate, which they suggest is the more accurate and appropriate one, the Bayes factor of $4x10^6$ indicates almost no chance the species might be extant. In summation, the hard evidence available to modelers casts serious doubts on the validity of the species' "rediscovery" in 2004[106], and finds little justification for the subsequent, costly search to find more conclusive evidence of the ivory-billed woodpecker's existence. Some argue the search continues as long as hope does, but statistics has a somewhat different answer in this case. And with other species like the Spix's macaw still potentially within the bounds of rescue, the resources of conservation organizations might be better devoted to saving those species than to chasing the ghosts of woodpeckers past.

Once it is determined that there is an acceptable level of probability that a species is extant, one possible way to further leverage the data collected would be use the data to build species distribution models (SDMs) to aid in the search and rescue effort. In basic terms, SDMs use information about the conditions where a species has occurred (and where it has not occurred) to determine the realized ecological niche of the species. This niche can be projected onto geographic space to help identify areas areas that appear highly suitable for the species but perhaps have not been searched yet [107]. This approach has been successful in identifying new populations of threatened species (e.g. [108]), with the author identifying new populations of four of the eight rare plant species in the study. While SDMs are commonly used in a variety of different ecological and conservation applications, there is a deep literature on comparisons of SDM methods (see Qiao *et al.*[109] for an overview), so caution must be exercised in selecting which methods are best for the available occurrence and environmental data. This approach—of determining the probability a species is still extant and using SDMs to identify the areas they are most likely to be—may provide a way forward for conservation agencies for making cost-effective decisions of which species to pursue and where to look for them.

2.6 The Community Scale and Beyond

Suppose that, in a twisted experiment motivated by an ecology-related childhood trauma, a mad scientist was developing a scheme to reduce global biodiversity to one half of the

Earth's total species. Hunting, fishing and poaching could achieve that goal slowly, but would be particularly inefficient for eradicating insects; and while a generalist disease might help eradicate a handful of mammals or a sizeable fraction of amphibians, the majority of species would still remain. But perhaps realizing that habitat loss might be the most efficient tool to destruction, that scientist might cut the Gordian knot, by simply bisecting the Earth and destroying half. Would his plan come to fruition?

Our mad scientist's plan is riddled with flaws. If half of the species were endemic to each half of the Earth with no overlap, his plan would succeed. But a handful of species in any clade of life are globally-cosmopolitan; and no matter how his plan was executed, the handful of species occurring on both halves of the Earth would leave him with far, far more than half the species he started with.

With renewed vigor, the mad scientist sets out on a newly ambitious project: what percentage scorched earth would be required to achieve his goal? He begins by counting every species on his sidewalk block, then in his neighborhood, and up to bigger scales. With enough grant funding and undergraduate assistants, he has eventually covered a measly 6.25% of the Earth when he realizes he has counted half of Earth's species. To enact his master plan, he's tasked with destroying the remaining 93.75%. Going by land area alone (his grudges, we suppose, do not extend to the ocean), he only needs preserve 3.6 million square miles of land - roughly (conveniently?) the land area of the United States.

The process our nationalist, isolationist villain has enacted is the empirical construction of the species-area relationship (SAR), one of the oldest and most powerful scaling laws in macroecology. Because the synthesis of different factors at global scales is challenging, and habitat loss is one of the easiest extinction drivers to measure, the SAR gives us a powerful tool for approximating extinction rates - at the price of not knowing specifically which species will go extinct.

The Species Area Relationship

The biogeographer Olof Arrhenius began the process of formalizing the SAR in a classic 1921 paper in the *Journal of Ecology* titled "Species and Area." [110] In it he observed that, by expanding the area of focus, the number of species continues to increase at a diminishing rate (but, never reaching an asymptote [111]). The canonical formula for the SAR has come to be called the Arrhenius SAR, and is formulated as

$$S = cA^z$$

where c is a constant fit to the data, and z is a slope conventionally treated as 0.25. The application of that formula to extinction rate estimation is relatively obvious: by changing the amount of area, we can change the number of species:

$$S' = c(A')^z$$

and calculate the number of extinctions

$$E(A') = S - S'$$

In our mad scientist's failed scheme, reducing the area of the Earth by half would leave us with far more than half the species:

$$\frac{S'}{S} = \left(\frac{0.5A}{A}\right)^{0.25} = (0.5)^{0.25} = 0.84$$

In a 2004 *Nature* paper that has become perhaps the most cited study on extinction since the millennium, a group of researchers led by Chris Thomas refined the global extinction rate estimate by analyzing species' habitat losses from climate change and applying the SAR. Their extinction-area relationship took three forms applied to n species, with a given A_i area per species before change, and A_i' subsequent to habitat loss:

$$E_1 = 1 - \left(\frac{\sum\limits_{i\in(1,n)} A_i'}{\sum\limits_{i\in(1,n)} A_i}\right)^{0.25}$$

$$E_2 = 1 - \left(\frac{1}{n}\sum_{i\in(1,n)} \frac{A_i'}{A_i}\right)^{0.25}$$

$$E_3 = \frac{1}{n}\sum_{i\in(1,n)} \left(1 - \left(\frac{A_i'}{A_i}\right)^{0.25}\right)$$

Using those three methods in combination with species distribution models, the authors estimated that 15-37% of species on Earth might face climate-driven extinction by 2050. That result is by far one of the most important ones produced in any study of extinction, and has supported a number of the most expansive conservation programs worldwide.

Refining and Reformulating the SAR

Like many "laws" of ecology, the conventional SAR has problems and pitfalls, and with the tremendous array of approaches developed to study it, it has even been called ecology's "most protean pattern."[112] Subsequent to the publication of Thomas *et al.*'s study, one of the most seminal debates in extinction research has centered around its conclusion that climate change is likely to act as the most consequential driver of the sixth mass extinction. Different approaches to the species area relationship, and comparable or derivative macroecological methods, have sprung up in the wake of Thomas's work. Here, we review a few of the different approaches that can be used to predict extinction rates at the community level.

z: a Dynamic Scaling Property

The most immediate problem with applying the species area relationship is that the slope z, normally set to 0.25, is neither universal nor scale-independent. In part, this is because of two different constructions of the SAR. The slope of 0.25 derives from the experimental work of Macarthur and Wilson on island ecosystems, which applied the SAR to the richness of species on islands of different sizes. For islands (and for application of the island SAR to extinction), a slope of 0.25 is justified under a set of three (relatively common) circumstances delineated by Harte and Kitzes: "(i) total abundance in the new area A is proportional to area, (ii) individuals found in A are chosen by a random draw from all individuals in A_0 [the initial area], and (iii) the number of individuals of each species in A_0 follows a canonical lognormal abundance distribution." [113]

However, the continental "nested" SAR (constructed from nested areas on a continental scale) does not always follow the same property. This is in part because the conventionally-used SAR assumes self-similarity (or in more tangible terms, picking two patches of different area always yields a roughly-the-same-slope difference in species). As it turns out, self-similarity works within some sites but not others, and within the Western Ghats mountains of India alone, scaling up from vegetation sampling plots to broader scales brings z down from values closer to 0.5, down to values approaching 0. [114] Selecting an appropriate slope based on scale is an important part of appropriate use of the SAR to predict extinction rates, and as analyses approach the continental scale, the appropriateness of the SAR method decreases as z approaches zero.

An Alternate Approach Based on the Endemics Area Relationship

In the Thomas *et al.* study, the application of the species area relationship followed three methods, and while some explicitly predicted extinction risk at the scale of a single species, all rely on the prediction of reduced species richness based on habitat loss. In place of this indirect calculation of decreased richness, a more direct approach uses what is called the *endemics area relationship*, which calculates the number of endemic species restricted to a given area (all of which should be committed to extinction when the area is destroyed). As pointed out by He and Hubbell, the SAR and the EAR are not mirror curves except in a single special case when species are completely randomly distributed in space; else, the "forwards" and "backwards" methods of extinction calculation are not, they argue, comparable. [115]

Prediction of extinction based on the EAR may be more appropriate for measuring the immediate effects of habitat loss, and is likely to better account for the "geometry of habitat clearing." [116] Storch *et al.* [117] developed an approach to the SAR and the EAR that scales the area by the mean geographic range size in the focal clade/area and scales richness by the average number of species in that mean geographic range. When plotted, the SAR curves upwards while the EAR is roughly linear with a slope of 1 across most scales. Starting from basic knowledge about the average geographic range size of a given species, this result indicates that extinction from habitat loss can be predicted based on the EAR across scales

fairly accurately.

An Alternate Approach Based on Maximum Entropy

Two "unifying" theories have dominated discussions about macroecology. The first is the unified neutral theory (UNT) of biogeography and ecology (proposed by Stephen Hubbell), which is beyond the scope of this chapter; the second is the maximum entropy theory of ecology (METE) proposed by John Harte. The METE deserves special mention here, due to a particular focus in the METE literature on improving the applicability of the SAR to extinction rate prediction. What differentiates both the UNT and the METE from more general conceptions of the SAR is the explicit treatment of species abundance as a component of community assembly. The theory of the METE is far too complex to encapsulate in this chapter (and an entire book by Harte exists for that purpose), but a few useful derivations are worth mentioning. One is the derivation by Kitzes and Harte of an extinction probability that is applicable at the species scale [113] based on proportional area loss (A_0/A, shortened to β) and corresponding reduction in abundance (n from n_0) with a general probability distribution

$$P(n|n_0, A_0, A) = ce^{-\lambda n}$$

for which they provide rough approximations

$$c \approx \frac{1}{(An_0/A_0) + 1}$$

and

$$\lambda \approx \ln\left(1 + \frac{A_0}{An_0}\right)$$

Drawing on similar concepts from the pseudoextinction thresholds we discuss above in §4.1, they suggest the probability a remainder $r_c = n/n_0$ will be left after habitat loss is equivalent to

$$\mathrm{P}rob\left[\frac{n}{n_0} > r_c\right] = \int_{r_c n_0}^{n_0} ce^{-\lambda n} dn = \frac{[n_0\beta/(1+n_0\beta)]^{r_c n_0} - [n_0\beta/(1+n_0\beta)]^{n_0}}{(1+n_0\beta)\ln(1+1/n_0\beta)}$$

Given a starting population and a critical population size, analogous results can be derived for the Thomas *et al.* calculations; and higher level predictions can be made based on the distribution of abundances and critical abundances within the community.

In a subsequent publication [118], this *extinction area relationship* is extended even further to extrapolate a MaxEnt-based probability that a given number of species will remain after habitat loss. It assumes a logseries distribution ϕ of abundance for species with a mean μ_ϕ with a single shape parameter p

$$\phi(n_0) = \frac{-p^{n_0}}{\ln(1-p)n_0}$$

$$\mu_\phi = \frac{-p}{(1-p)\ln(1-p)}$$

They similarly propose an upper-truncated geometric species specific abundance distribution, which provides the probability n individuals remain in a fractionally reduced area a (β in their other notation) based on a shape parameter q:

$$\Pi(n|a, n_0) = \frac{(1-q)q^n}{1-q^{n_0+1}}$$

where q is solved implicitly based on a and n_0 from

$$an_0 = \frac{q}{1-q} - \frac{(n_0+1)q^{n_0+1}}{1-q^{n_0+1}}$$

The probability a species is found in area A after habitat loss follows a distribution g that takes the form

$$g(a, n_c) = \sum_{n_0=1}^{\infty} (1 - \Pi(n \leq n_c|a, n_0))\phi(n_0)$$

which scales up to a community-level richness after area loss

$$p(S|S_0, g) = \binom{S_0}{S} g^S (1-g)^{S_0-S}$$

where

$$g(a, n_c, \mu_\phi) = \sum_{n_0=1}^{\infty} \left(1 - \frac{q^{n_c+1}-1}{q^{n_0+1}-1}\right)\left(\frac{-p^{n_0}}{n_0 \ln(1-p)}\right)$$

or if the pseudoextinction threshold is set to zero (i.e. no species has 0% survival odds until all individuals are dead) and area loss is severe, that expression can be reduced to eliminate the q term:

$$g(a, n_c, \mu_\phi) = -\frac{a}{\ln(1-p)} \sum_{n_0=1}^{\infty} \frac{p^{n_0}}{an_0+1}$$

This METE approach thus provides a *probabilistic species area relationship* (PSAR) that can be used to provide not only an expected extinction rate under habitat loss but also a range of confidence. This becomes an especially important tool in a small community of only a few dozen species or fewer (or in communities with pervasive low abundance across species), where deviations from SAR-based predictions may be greater due to stochastic processes.

How does the PSAR scale up against the Thomas-SAR? It has a clear advantage in the prediction of individual species extinction risk (but correspondingly requires more data on abundance/demography that may be absent for many poorly-known taxa). Kitzes and Harte provide two illustrations; first, assuming the normal slope of 0.25, the PSAR predicts a 44% chance of extinction for a species that loses 90% of its habitat. Second, if we assume a pseudoextinction threshold of 50 individuals, by comparison to the predictions of the PSAR, the Thomas-SAR approach underpredicts extinction risk if $n_0 \leq 1000$ but overpredicts if n_0 is greater–supporting the notion that the 15-37% extinction rates that Thomas *et al.* study predicted could be an overestimate.

Tying Up Loose Threads, Thinking Across Scales

The various different approaches to predicting extinction at the broadest scales have driven substantial controversy among different interpretations of macroecological theory. But one of the most important problems is that estimates of extinction from these methods are still poorly connected, by and large, to the rest of the extinction literature, and to the other types of models we discuss above. One of the most innovative and unusual approaches in the literature was presented by Rybicki and Hanski [119], who simulated a stochastic patch occupancy model (similar to those presented in §3.1) with spatially heterogeneous environmental conditions across patches. While their model incorporates the standard mainstays of an SPOM (colonization, extinction, a dispersal kernel), it also incorporates a phenotype and niche breadth that produce a Gaussian fitness function (much like the models in §4.1)

Tying together a number of the important ideas discussed above, the work of Rybicki and Hanksi made several advances into new territory. For one, they make a semantic distinction between the endemics-area relationship (EAR, which they define as the $S = cA^z$ relationship applied to the area lost $A = a$) and the "remaining species-area relationship" (RAR),

$$S - S_{loss} \;=\; c(A_{new}/A)^z$$

The EAR and RAR, as two methods of calculating extinction risk, are not interchangeable or symmetric counterparts. Rybicki and Hanskii highlight a discrepancy between Storch *et al.*'s suggested EAR slope of roughly 1, and He and Hubbell's values which were a tenth smaller,[115] which they suggest can be resolved by the fact that Storch fit the EAR while He and Hubbell were calculating the RAR; and their simulations agree with the results of He and Hubbell that the slope of the RAR may be half or less that of the SAR.

Their empirical approach to simulation leads to a valuable conclusion that stands in opposition to previous work. While Kinzig and Harte[116] and He and Hubbell [115] both strongly suggest that the SAR overestimates extinction risk; the results of Rybicki and Hanskii's simulations suggest that in the short term, the RAR underestimates extinction while the continental SAR ($z \approx 0.1$) is adequate. Their result ties the population scale to the community scale, as they attribute it to species' populations *outside* destroyed or fragmented habitat falling below critical thresholds and facing extinction despite the lack of total endemic extirpation. In the long term, they suggest, the island SAR ($z = 0.25$) may

be the best predictor of total losses. Finally, they explore the difference between leaving a single patch of habitat and fragmenting habitat, and conclude all models underestimate extinction risk in scenarios of extreme fragmentation. To address that problem, they propose a modified species area relationship

$$S = cA^z e^{-b/\lambda_M}$$

where λ_M is the metapopulation capacity (see §3.1) and b is another scaling parameter like c and z. If n is the number of habitat fragments, they suggest, the metapopulation capacity scales linearly with A^3/n^2, meaning that the *fragmented landscape species area relationship* (FL-SAR) can be expressed as:

$$S_{new}/S = (A_{new}/A)^2 e^{-bn^2/A^3}$$

While the data to fit such an expression might be challenging to collect (and so the FL-SAR may not be an immediately useful conservation planning tool), the FL-SAR provides an important and much needed link between the population and metapopulation processes we discuss above, and our broader understanding of the rate of extinction at landscape and community scales.

2.7 Last Chance to See

What don't we know about extinction yet?

As predictive tools gain precision, our estimates of the extinction rates of well-known groups like mammals and birds also become more precise. But the majority of the world's species are not yet known; most animal diversity is harbored by insects or parasites (especially nematodes), and the vast majority of species in those groups are undiscovered or undescribed. Their extinction rates are just as poorly quantified as their diversity or the hotspots of their biodiversity. But some basic estimates suggest that 7% of the planet's invertebrates may have already gone extinct—at which rate evidence would suggest that 98% of extinctions on Earth are currently going undetected. [120] It's also especially difficult to compare these extinction rates to historical baselines, because the fossil record for most invertebrates and other taxa are incomplete or nearly absent.

An especially poignant problem is the detection and estimation of coextinction rates—the secondary extinction of species dependent on others for their ecological niche—which Jared Diamond suggested in 1989 was one of the four horsemen of mass extinction (in his words, "overhunting, effects of introduced species, habitat destruction, and secondary ripple effects"). [121] Among the most obvious candidates for coextinction are two main groups: pollinators (which can have a strict dependency on host plants) and endosymbionts (parasites and mutualists, which may exhibit strict specificity in their association with plant or animal hosts). While both groups are believed to be severely at risk of secondary extinction, quantifying their extinction rate can be challenging, as there is rarely a 1:1 correspondence

between hosts and dependent species. An approach popularized by Koh simulates host extinctions in a random order and predicts the number of corresponding coextinctions from the *affiliation matrix*; by fitting a function to real affiliation matrices, Koh *et al.* found that if host specificity is 1:1 then the slope is linear, but when affiliates use a greater number of hosts, the coextinction function is concave upward:

$$\bar{A} = (0.35\bar{E} - 0.43)\bar{E}\ln\bar{s} + \bar{E}$$

where E gives primary extinction risk, A gives secondary extinction risk, and s is host specificity [122]. Subsequent work has shown that even though parasites and mutualists may experience a reduced rate of extinction from host switching, the majority of threatened species on Earth might still be mutualists and parasites (due to the tremendous diversity of such species, e.g. the estimated 300,000 species of helminth alone). [3] One recent study using the Thomas species-area relationship approach estimated that, from the synergistic pressures of climate change and coextinction, up to one third of all helminth parasites might be threatened with extinction by 2070. [123]

Most affiliate extinctions are poorly cataloged, if recorded at all [124], and only limited conservation frameworks exist for their study. More data is needed on host-symbiont association networks to better inform the role that non-random structure in those networks might play in increasing or decreasing extinction rates; some work has suggested that species preferentially favor more stable host species, the underlying cause of a "paradox of missing coextinctions." [125] Similarly, the potential for species to switch hosts and thereby avoid extinction is unknown, but likely mitigates global extinction risk. In parasitology, the Stockholm Paradigm suggests that host-parasite associations diversify in changing climates and environments as a function of (1) phenotypic plasticity, (2) trait integration and (3) phylogenetic conservatism of "latent potential," which together produce a pattern of *ecological fitting* that might benefit parasites (and thereby other symbionts) in the face of the sixth mass extinction. [126] A more in-depth treatment of the theoretical ecology of ecological fitting can be found in the recent work of Araujo *et al.* [127]

Is saving microbes and parasites from extinction a reasonable goal? Some argue that it is [128], but others have recently suggested it's "time to get real about conservation" and focus on our failure to adequately prevent catastrophic population crashes in megafauna like elephants [129]. Regardless of animal type or conservation status, the development of demographic theory and predictive modeling are our best options to understand and mitigate extinction risk in natural populations. One such advance deserving of special mention is the development of *early warning signals* of population collapse. This is a developing body of literature that is built around the fact that populations on the verge of collapse often produce detectable statistical signals [130] that, detected far enough in advance, might allow mitigation efforts and prevention of population collapse.

The majority of early warning signals for extinction currently rely on identifying *critical slowing down*, a process by which as the dominant eigenvalue of the system tends towards zero, populations return to equilibriums lower after perturbations, with increasing variance

and temporal autocorrelation. Critical slowing down is often a sign of a dynamical system approaching a bifurcation point, which may sometimes indicate a non-catastrophic shift to an alternative stable state [131], but more usefully, may indicate an impending extinction emerging from a shift into a sharp decline or into chaos. A foundational experiment by Drake & Griffen showed that critical slowing down can readily be detected from time series data for populations facing decreasing food availability up to eight generations before they reach the bifurcation point [132]. Conventionally, this is done by taking a set of metrics such as the autocorrelation, coefficient of variation, or skewness (termed leading indicators), scaling and adding them into a single metric (composite indicators), and tracking them over time. Given any indicator w, the standardized statistic is calculated as a function of the mean and standard deviation of the time series to that point:

$$\hat{w}_t = \frac{w_t - \bar{w}_{1:t}}{\sigma(w_{1:t})}$$

When the test statistic $\hat{w}$ passes a threshold level of deviation from the running average $\bar{w}_{1:t}$, such as 2σ, this can be taken as an early warning signal. [133]

Early warning signals are far from a perfect tool. Most research has focused on detecting critical slowing down, but not all types of dynamical systems exhibit critical slowing down [134, 135]. Even if critical slowing down is expected, these types of early warning signals are far from perfectly accurate. Ecological data, especially from population abundance estimates, often has a high signal:noise ratio [134], to which early warning signals are still sensitive, leading to an expected mix of both Type I and II errors, depending on the quality and quantity of data. In some cases, this problem can be accommodated for by evaluating early warning signals as an iterative process over the time series, rather than taking the first warning as the only required evidence. [136]

Simulation work has shown early warning signals to be fairly robust to incomplete sampling, but proportionally data intensive to a degree that may be impossible to reach with available ecological time series data [137]. When population data is lacking or incomplete, trait-based data can be used as a proxy or a supplement, if traits such as body size are expected to correlate with fitness, and are responsive to ecological shifts. [133] One particularly interesting demonstration by Clements *et al.* showed that body size data could be used to predict the collapse of whale populations 10-40 years before the whaling industry depleted stocks. [136] Other cutting-edge work is attempting to scale the detection of early warning signals to the metapopulation level by developing spatial early warning signals [131], which could be used to optimize reserve design and address the influence of dispersal, stochasticity, and local population dynamics on metapopulation persistence. Some work has even suggested critical slowing down could be used to identify tipping points of network collapse due to serial coextinctions in symbiont networks. [138]

The pressure for more accurate, predictive tools will only grow in the next few decades of research. A recent review by Mark Urban surveyed studies of climate change-driven extinction risk and found that, despite the variation between different modeling methods and scopes, projected extinction rates are not only rising but one in six species might be

imminently threatened with extinction. [139] Similarly, in a study of roughly 1000 species of plants and animals, about half had experienced population extinctions driven by climate change. [140] As extinction rates accelerate due to global change and we fully enter the sixth mass extinction, the need for better analytical and simulation tools—that produce precise estimates from limited data—will only grow. In light of the constant need to test, revise and re-test models of extinction, to a mathematically-trained ecologist or an ecologically-minded mathematician, this field of research is a critical opportunity to apply the principles of ecosystem science towards a high-impact and worthy goal.

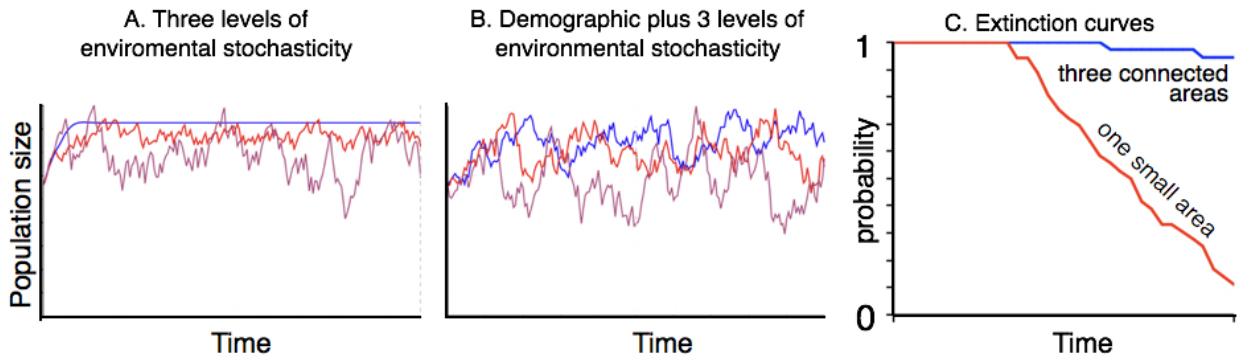


Figure 2.1: An example PVA without (A) and with (B) the influence of demographic stochasticity, and with no (blue), medium (red) or high (purple) environmental stochasticity. Based on many numerical simulations, an "extinction curve" can be plotted from the probability of population survival over time (C). This analysis can be used to make decisions about management and conservation: here, illustrating that three populations with migration between them survive for much longer in a poached population of rhinos than a single population. An interactive tutorial of PVA, which can be adjusted to produce anything from the simplest population dynamics to a stochastic, structured metapopulation experiencing harvesting, can be found at http://www.numerusinc.com/webapps/pva.

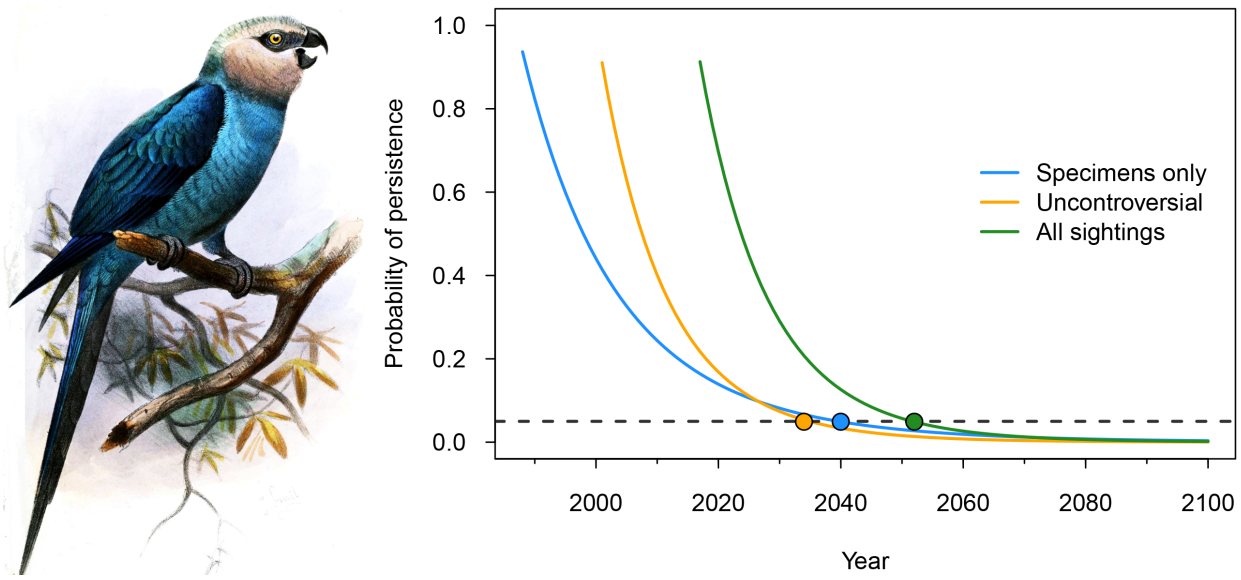


Figure 2.2: Estimates of likely extinction date of the Spix's macaw based on extinction estimating equations in [91]. The lines represent the estimated probability the species is extant each year; the blue line is the results using physical evidence only (specimens / wild-caught individuals), the orange line for uncontroversial sightings *and* physical evidence, and the green line is the results for all sightings, including controversial. The dotted line is a significance level of 0.05. Once the probability drops below this level, the species is considered likely extinct.

Chapter 3

Parasite biodiversity faces extinction and redistribution in a changing climate

Colin J. Carlson Kevin Burgio Eric Dougherty Anna Phillips Veronica Bueno Christopher Clements Giovanni Castaldo Tad Dallas Carrie Cizauskas Graeme Cumming Jorge Doña Nyeema Harris Roger Jovani Sergey Mironov Oliver Muellerklein Heather Proctor Wayne Getz

3.1 Abstract

Climate change is a well-documented driver of both wildlife extinction and disease emergence, but the negative impacts of climate change on parasite diversity are undocumented. We compiled the most comprehensive spatially-explicit dataset available for parasites, projected range shifts in a changing climate, and estimated extinction rates for eight major parasite clades. Based on 53,133 occurrences capturing the geographic ranges of 457 parasite species, conservative model projections suggest 5-10% of these species are committed to extinction by 2070 from climate-driven habitat loss alone. We find no evidence that parasites with zoonotic potential have a significantly higher potential to gain range in a changing climate, but we do find that ectoparasites (especially ticks) fare disproportionately worse than endoparasites. Accounting for host-driven coextinctions, models predict that up to 30% of parasitic worms are committed to extinction, driven by a combination of direct and indirect pressures. Despite high local extinction rates, parasite richness could still increase by an order of magnitude in some places, as species successfully tracking climate change invade temperate ecosystems and replace native species with unpredictable ecological consequences.

3.2 Introduction

The biotic footprint of global climate change, quantified in the shifting distributions and extinctions of animal and plant taxa, has been a subject of intense research since the turn of the century. [141, 142, 19] While some species can track shifting climates[143], many likely face extinction at rates projected a decade ago to be as high as 15-37%. [144, 8] Despite recent refinements of that estimate finding the overall rate of climate-change-driven extinction is likely closer to 8% [139], others suggest that if current extinction rates (from climate change and other anthropogenic impacts) persist for hundreds to thousands of years, total extinctions could cross the 75% threshold that defines a geological mass extinction event. [20] However, previous work has focused nearly exclusively on free-living biodiversity (especially vertebrates), and many important functional or taxonomic groups remain undescribed or are only now being included in extinction research. [145] Particularly poorly profiled are commensalists, mutualists, and parasites [146, 147], which should exhibit an atypically high extinction rate due to their dependence on other species for survival. [124]

Despite substantial research on parasite coextinction risk [124, 146] and an emerging body of theoretical work predicting the potential adverse impacts of climate change on parasites, climate-change-driven extinction rates have never been estimated for parasitic groups, perhaps because the long-term data needed to detect extinctions-in-progress had not been previously collated. [148] A recent study predicts that one of the most reliable benchmarks of parasite extinction risk should be their loss of suitable habitat, but notes that distributional data are lacking for most parasites. [1] For species with available data, two frequently cited studies contrastingly predict either local range loss [149] or global range increases for ixodid ticks. [150] Even for zoonotic parasites, which are closely monitored compared to the majority of parasitic organisms on Earth (especially in the context of climate change research [1]), the net relationship between climate change and disease emergence is uncertain.[151]. Early work argued that a warming climate facilitates range expansion [152] , though others have predicted that range shifts will be accompanied by little expansion [153]. Further evidence suggests that some zoonoses–like the nematode that causes angiostrongyliasis in humans–could lose suitable habitat as a result of climatic changes. [154]

If parasites face severe extinction risk in a changing climate, the cascading impacts on ecosystems are likely to be profound. Many parasites play an important immunoregulatory role in host populations, and some studies have found that a higher diversity of parasites can act as a partial buffer against the emergence of a virulent pathogen. [10] Previous work has also pointed to the merits of parasites as regulators and connectors in resource-consumer webs, in which they can sometimes constitute more than 75% of the total links, and in which their occasional role in altering host behavior can be critical to the flow of biomass between trophic levels. [5] Despite their many hidden benefits, parasites are a difficult subject for conservation research, as parasites can come at an economic (e.g., crop pathogens) or health (e.g., emerging infectious diseases) cost to wildlife and human populations. In the context of climate change research, the balance between parasite extinction and emergence is uncertain, and while some work has suggested these could be complementary processes [155], the net

impact of climate change on parasite biodiversity is still unresolved.

Here we use an updated and improved set of climate envelope models, conceptually based on those of Thomas *et al.* [8], and apply them to a dataset we have assembled that contains the most comprehensive multi-species occurrence dataset available for clades containing terrestrial macroparasites or groups historically treated as parasites, all of which have often been overlooked in extinction risk estimation. We focus on eight major clades: Acanthocephala (spiny-headed worms), Astigmata (two superfamilies of primarily ectosymbiont feather mites), Cestoda (tapeworms), Ixodida (ticks), Nematoda (roundworms), Phthiraptera (lice), Siphonaptera (fleas), and Trematoda (flukes). Our data combine and refine existing online repositories and newly digitized museum collections. Over 170,000 unique candidate points were reduced through strict data cleaning, quality control, and sample size limits to 53,133 georeferenced parasite occurrence records, cataloging the distributions of 457 species (Figure 3.1). From these data, we constructed maximum entropy species distribution models, and projected each species' distributional shift for the year 2070 under an ensemble of climate change models and scenarios (Figure 3.2). As in the original Thomas *et al.* study, we forecast loss of "native" range (i.e., loss of areas currently occupied: "0% dispersal") and compare it to overall changes in suitable range (including in areas not previously occupied: "100% dispersal", "global"). The resulting maps of current and projected ranges, forecasts of extinction risk and habitat loss, measures of zoonotic potential, and model accuracy metrics are available through the Parasite Extinction Assessment & Red List (PEARL), an atlas that is available online at pearl.berkeley.edu (Figure 3.3); and the data are available as supplementary files on request.

3.3 Materials and Methods

Data collection and georeferencing.

Accurately describing the distribution of global parasite biodiversity is a nearly insurmountable task, and for a study of this nature, might be essentially impossible without taking advantage of existing data infrastructure, especially from natural history collections [156]. Globally, we project extinction risk using a patchwork of regionally and taxonomically specialized datasets representing the best available distributional datasets in parasitology, including: a published database on ixodid ticks in Africa [157, 158], a comprehensive database of all published records of feather mite occurrences recently published as a data paper in Ecology [159], data from the Global Biodiversity Informatics Facility (www.gbif.org), the University of Michigan Museum of Zoology (UMMZ)'s database of bee mite occurrences, flea data from the VectorMap project (vectormap.si.edu), and most significantly, the U.S. National Parasite Collection (USNPC) (Figure 3.1). The datasets included in this study represent some of the only true "big data" for wildlife parasites; a number of the largest datasets do not include any spatially-explicit data, such as the London Natural History Museum database [160] and the FishPEST database. [161] Others, like the Global Mammal

Parasite Database, have begun georeferencing but data are not public (P. Stephens, *pers. comm.*); and some sources, like the University of Connecticut's Global Cestode Database, do not have spatially-explicit data that met the sample size requirements delineated below. [162] Finally, a set of researchers noted in the acknowledgements of the main paper generously donated their data, but geographically-biased focus led to their exclusion from the final dataset.

In cases within these databases where occurrences lacked coordinate data, we georeferenced all specimen locality data using established guidelines in the literature [163] and the georeferencing software GEOLocate. [164] For the remaining databases with existing georeferenced occurrences, we largely used the data in its original form (i.e., the Cumming tick database [33,989 points] and UMMZ bee mite dataset [1,160 points]). The VectorMap dataset includes mites, ticks, and fleas, but the mite data did not include any species with 20 or more occurrences, and the ticks were identical to GBIF and Cumming datasets, so only fleas were included adding an additional 8,482 points to the dataset. GBIF data was downloaded for eight clades (Acanthocephala, Astigmatina [= Astigmata], Cestoda, Ixodida, Nematoda, Phthiraptera, Siphonaptera, and Trematoda), and clipped to terrestrial points only. A number of other parasitic clades were also considered in the process of data collection, but GBIF data for the vast majority was limited and provided ten or fewer suitable species. These required specific cleaning to remove country-level datasets that would have produced biased niche models from data limited by administrative boundaries [165]—for example, the Edaphobase dataset and Dieter Sturhan's DNST dataset, both on nematodes in Germany. Records without species identification were eliminated, as were records without any coordinates. A total of 100,295 occurrences remained after the application of these guidelines (Acanthocephala: 2,013; Astigmata: 4,826; Cestoda: 3,048; Ixodida: 17,695; Nematoda: 32,170; Phthiraptera: 3,675; Siphonaptera: 23,573; and Trematoda: 13,294).

The georeferencing web-based software GEOLocate [164] was used to assemble a spatially-explicit database for the U.S. National Parasite Collection and for the feather mite database. While the USNPC has over 70,000 records, only some have locality data digitized, and of those, many have single or double records. To consolidate our efforts, we only georeferenced data for species with twenty or more records, a threshold chosen based both on the distribution of sample sizes within the larger dataset, and from literature evidence about the effect of sample size on model accuracy. While accuracy consistently stabilizes around 50 points for most commonly used methods [166, 167, 168], 20-or-more is often used as a threshold in the literature especially for MaxEnt (which particularly excels with small sample sizes), and a recent publication shows that 15 points often suffices for narrow-ranged species, and as few as 25 are sufficient for the most globally distributed species. [169] The 20-or-more rule reduced the dataset down to 31,212 specimens many of which had shared locality information (i.e., multiple species were collected from the same locality, often by the same collector). In the supplement, we present an additional analysis that only uses species with 50 or more points (Appendix 3), reducing the dataset from 457 species down to 196. As mentioned in the Results section, the main analyses are essentially unaltered by the reduction, with some extinction estimators producing slightly higher values (especially for

clades with already limited sample sizes). Consequently, in the main text, we have elected to present the most inclusive analysis, both maximizing the species coverage of our study, and producing a slightly more conservative estimate of extinction rates.

Points attributed to townships and provinces were marked at the political center and the uncertainty radius was the minimum to encompass the entire region. When political markers were not available for countries in the developing world, we used satellite imagery to identify the rough boundaries of cities and townships. In select cases, if the political marker was too far from the actual center, our final point would be corrected to the center of the region of uncertainty. Broad geographical regions like Siberia, entire stretches of river or continental coasts, and points with too many candidate options of equal likelihood (e.g., "Red Mountain, California, United States" which could be 17 localities, or "La Junta, Mexico", which had 33 candidates) were all excluded from the final dataset.

A total of 5,507 specimens from the USNPC were skipped by virtue of incomplete information, or because the locality data was insufficiently detailed, leaving 25,705 specimens georeferenced to 7,373 unique localities. Of those remaining entries, the data were reduced to species with enough occurrence points for inclusion. A final verification against the Smithsonian's collection management database revealed 21 inaccuracies that were subsequently corrected, and one nematomorph and five botflies were removed from the dataset, as were two monogeneans that had been formerly recorded as trematodes. The final dataset contained 15,741 unique entries for species with at least 20 points.

Uncertainty greater than 10 km has been shown to potentially negatively impact the accuracy of distribution models. [170] However, we used climate data with a resolution of 10 arc-minutes for our models, roughly an average of 20 km (variable due to the shape of the earth), a coarser resolution than many niche modeling efforts, and one that absorbed much of the error associated with the data. Consequently, we set a threshold of 40 km (two cells) for acceptable maximum uncertainty in our georeferenced points, and subsequently reapplied the 20-or-more-points threshold, eliminating some species in the process. Geolocation data for the feather mites dataset was collected using the same methodology. From each dataset, we removed all occurrence data that did not meet our criteria for a 40km uncertainty radius. We then excluded all species that did not have at least 20 high quality occurrence points.

In the final dataset aggregated across every data source, we provide results for a total of 457 species across the compiled datasets with a total of unique 53,133 points (an average of 116 unique points per species, far above the 20-point minimum we set and on par with some of the more comprehensive distribution modeling published for individual species in the literature; Figure 3.3). Of those 457 species, these were broken down as: Acanthocephala: 14 species; Astigmata: 18 species; Cestoda: 25 species; Ixodida: 141 species; Nematoda: 147 species; Phthiraptera: 5 species; Siphonaptera: 67 species; Trematoda: 40 species. Models are presented for all but one species in that set, as models ran unsuccessfully in that case, but locality data are still made available (the nematode *Teratocephalus terrestris*).

Climate data.

Distribution models were run using the WorldClim v1.0 climate dataset at 10 arc-minute resolution, which includes 19 bioclimatic variables (Bioclim) that capture global trends and variability in precipitation and temperature. [171] Based on models that have been previously published that forecast species distributional shifts, we selected a set of five of the most widely used global climate models (GCMs) at four representative concentration pathways (RCPs) that account for different global responses to mitigate climate change. These are the Beijing Climate Center Climate System Model (BCC-CSM1.1); Commonwealth Scientific and Industrial Research Organization, Australia's GCM (ACCESS-1.0); the Hadley GCM (HadGEM2-CC and HadGEM2-ES); and the National Center for Atmospheric Research's Community Climate System Model (CCSM4). Each of these can be represented at RCP2.6, RCP4.5, RCP6.0 and RCP8.5, capturing a range of scenarios species could experience (except ACCESS-1.0 which only exists for 4.5 and 8.5). Covariance between predictors and the definition of the accessible area are both significant problems with environmental predictors used in niche modeling efforts. To address these issues, models were trained on data subset to continents with known parasite occurrences, and regularization procedures in MaxEnt accounted for collinearity in predictor variables (see below).

Distribution modeling.

Our study was designed to reproduce the approach underlying the seminal Thomas *et al.* paper in Nature from 2004 [8], which projected that 15-37% of terrestrial species likely faced imminent extinction from climate change. [8] The overall order of operations is conserved: 1) "climate envelope" models are constructed using current best practices in ecological niche modeling; 2) species range shifts are forecasted in response to climate change; and 3) macroecological inference is made with respect to the consequences of that habitat loss for species extinction rates.

Since the publication of Thomas *et al.*'s study, the climate envelope method for estimating extinction rates has drawn some criticism. While climate envelopes (now more commonly termed ecological niche models, or species distribution models) are one of the most commonly used statistical methods in ecology, and hundreds of papers and several books outline best practices for their implementation [172, 173, 9], many researchers are still skeptical of the methodology. Niche concepts (foundational to climate envelope models) are contentious in ecology, and significant literature has been devoted over the years to basic assessment and debate about the utility and validity of niche theory as an approach to ecology. [174, 175, 176] More practically, climate envelope models fail to account for the role biotic interactions, source-sink dynamics, and dispersal play in determining range shifts. [177] Moreover, model performance can be challenging to accurately and honestly assess, especially given the non-independence inherent in the collection of many occurrence datasets. [127]

However, for species that are poorly documented in situ, the climate envelope approach remains popular as a tool for inference regarding geographic distributions. Parasites es-

pecially qualify for such an approach, given that mounting a field survey on the scale of our study would be nearly impossible, compared to the efficiency with which existing data sources can be used in the Thomas *et al.* framework. More broadly, in the absence of data that could parameterize more detailed methods like integral projection models, climate envelope models are still widely regarded as a powerful and popular method for studying climate change impacts on species ranges. Empirical work has shown even that climate envelopes validate well against real range shifts, especially as a "first-order approximation." [178, 179]

Where the Thomas *et al.* study is outdated from 12 years of updated scientific literature, we have correspondingly updated the methodology to improve accuracy and theoretical validity. Foundational to that methodology is the assumption that most or all species have a fundamental niche, measurable in multidimensional climate space, which constrains their geography. For parasitic species the relationship between bioclimatic variables and geography may be less intuitive than for plants or most animals, and in some cases, parasites with a high R0 and low environmental sensitivity may have ranges predominantly driven by their hosts. However, numerous cases exist where parasites are directly sensitive to climate. Among the many examples, humidity and aridity define geographic boundaries between feather lice on a shared host [180]; exposure to salt spray [181], altitude, and extreme cold negatively affect feather mites [182, 183]; precipitation and soil type can have a profound effect on free-living stages of helminths [184]; and further diverse support for parasitic niches independent of hosts comes from ticks [185], the plague bacterium (*Yersinia pestis*) [186], and even parasitic mistletoe. [187]

In response to Thuiller *et al.*'s [188] critique highlighting the significance of modeling method on forecasting range loss in response to the Thomas et al. paper in 2004, we have replaced the BIOCLIM algorithm (more commonly referred to in the literature now as "surface range envelopes") with maximum entropy regression, as developed by Phillips *et al.* [189, 190] and refined over the past decade. Maximum entropy (MaxEnt) is widely considered to be one of the best performing non-ensemble approaches to ecological niche modeling [191, 167], and is the most widely used method for the analysis of presence-only data in the literature as of right now. [192] MaxEnt models are frequently used to forecast species range shifts in response to climate change [193, 194], and have been found to often successfully predict the realized shift in species distributions. [195] However, MaxEnt allows fitting with up to five feature classes simultaneously (linear/L, quadratic/Q, hinge/H, product/P and threshold/T), and without careful tuning, has the propensity to overfit models more severely than many other comparable methods. [196] Consequently, using MaxEnt effectively relies on an approach that is sensitive to some major methodological pitfalls.

Sampling bias is a key problem in MaxEnt studies and correction for it can vastly improve predictive performance [197, 192], but some of the most direct solutions like spatial bias filters rely on knowledge that would be inconsistent across 457 species aggregated from different sources, with different relative biases. However, other approaches like cross-validation within models, feature class and variable set reduction (to reduce unnecessary complexity), and adjustment of the regularization parameter have all been shown to vastly strengthen MaxEnt models. [198, 196, 199] These processes are all automated in ENMeval,

an R package that performs cross-validation and model tuning [200]. We used ENMeval to automate analyses for all 457 species, with selection of the regularization parameter and feature classes based on an analysis using AICc, a version of the Akaike information criterion that is optimized for smaller sample sizes. Models with AICc of two or lower are considered to be "strongly supported" [201]; our automated process selected the model with the lowest AICc (AICc = 0), though alternative selections minimally affect the net results (Figure 3.4). This process substantially penalizes overfitting and overly-complex models, and as variable set reduction is also automated within the MaxEnt process, similarly reduces problems arising from covariance between Bioclim variables. Cross-validation was performed using the "checkerboard2" method which executes geographic cross-validation at a relatively broad scale, helping to account for some degree of sampling bias within datasets. While the jackknifing approach is more strongly recommended for species with small sample sizes (under 20-25 points [202]), our dataset already excluded species with small sample sizes by this definition, and having a single standardized model selection method across all species likely reduced the amount of noise in the final results (that is to say, differences between species are more likely to predict different outcomes under climate change, rather than inconsistency in methods).

For every analysis at the species level, a PDF is available on request that plots AICc, mean AUC (area under the receiver-operator characteristic curve), and training-minus-test AUC difference against the regularization multiplier , for six considered feature class sets: L, Q, H, LQH, LQHP, and LQHTP. These tuning aspects had no consistent effect on habitat loss projections (Figure 3.4). In cases where two models had the same AICc because features were included but not used, the more minimal-feature model (of the identical pair) was selected. The AUC and true skill statistic (TSS) were calculated subsequently for each species, and due to problems inherent in metric inflation with AUC that have been previously discussed in the niche modeling literature [203, 204], especially for datasets with spatial bias in their collection [205], we instead rely on TSS to measure final model validity. In ensemble approaches, modelers often exclude runs under a certain TSS to maintain quality, with values ranging from 0.3 to 0.85. [206, 207] According to Coetzee *et al.*, interpretation of the TSS can be roughly conceived of as "values from 0.2 to 0.5 [are] poor, values from 0.6 to 0.8 [are] useful, and values larger than 0.8 [are] good to excellent." [208] Based on that and other work using an 0.6 threshold [209], we present a comparison of habitat loss for all models and for those with TSS > 0.6 (Figure 3.4, Figure 3.5), noting that the pattern is essentially unchanged. However, all models of range shifts are made available on the PEARL server with AUC and TSS information presented alongside the models, to improve transparency and allow further work the option to be more selective.

Final models, selected based on lowest AICc, are projected onto our set of 18 GCM/RCP combinations (Figure 3.4). Non-logistic outputs of current and future projections are cut off by a threshold chosen to simultaneously maximize sensitivity and specificity (i.e., maximizing the TSS), turning outputs into binary geographic ranges. Differences in area between the two are most easily analyzed in R by using the BIOMOD_RangeSize function from the BIOMOD2 package. Differences in suitable area are calculated with 0% and 100% dispersal

(i.e., change in "native" and "global" habitat; Figure 3.5).

Likely outcomes between 0% and 100% dispersal will be a product of a species' own dispersal ability and the dispersal ability of their hosts. Hosts at the leading edge of range shifts may escape some parasitic infection [210, 211], but in the case of some ectoparasites, their ranges may be less constrained by host distributions. [185] Further analyses meant to better optimize accuracy and discriminate extinction risk between species should likely simulate the simultaneous shifts of host and parasite ranges, and thereby refine intermediate forecast scenarios for habitat loss. These will likely produce less conservative estimates of habitat loss, based on evidence that every life cycle stage can compound host-parasite spatial mismatch when climate change drives range shifts [212], though for some generalist groups like ticks, hosts may be a minimal constraint on geographic range size. [185]

Extinction rate estimation.

Distributional shifts modeled for parasites in each regional dataset were used as input to the same three species-area curves given by Thomas *et al.* [8] to estimate regional extinction risk:

$$E_1 = 1 - \left(\frac{\sum A_{new}}{\sum A_{old}}\right)^{0.25} \tag{3.1}$$

$$E_2 = 1 - \left(\frac{1}{n_{species}} \sum \frac{A_{new}}{A_{old}}\right)^{0.25} \tag{3.2}$$

$$E_3 = \frac{1}{n_{species}} \sum \left[1 - \left(\frac{A_{new}}{A_{old}}\right)^{0.25}\right] \tag{3.3}$$

The species area relationship with an exponent of 0.25 has come under methodological criticism by authors like Harte [213] and Kitzes [118] among others [115] (see also Appendix 2), but remains a widely used method in the literature from the last 1-2 years, often in combination with a similar RCP-structured approach to the one we use here. [214] We retain the method with the disclaimer that it is, in the current literature, more an "index of extinction" than a quantitatively strict prediction. We further suggest future studies investigate the applicability of dynamically-scaling SAR methods [114], which require data on population size and aggregation that are unavailable for most, if not all, parasites in this study. Calculation of compound risk was done using extinction estimates with and without dispersal, in combination with the coextinction risks presented in Appendix 1. This was done by assuming the two probabilities—direct extinction due directly to climate change and coextinction due to extinction of hosts—to be independent thereby yielding:

$$Prob[extinction] = 1 - (1 - p_{direct}) * (1 - p_{coextinction}) \tag{3.4}$$

Numbers presented in the main text are calculated as the product of values from Table 3.1 and Appendix 3.1 for the four main worm clades in our study.

Species richness mapping

A global map of species richness was constructed by stacking each distribution model included in the study, and counting the total number of species predicted to be present in each cell. We also present the turnover in each cell, like many previously published studies of climate change impacts. [215, 216] However, the role of sampling bias in the spatial structure of species richness cannot be overlooked; we devised an analysis based loosely on the quantile subtraction method employed by Hopkins & Nunn for gap analysis ([217]) to correct for sampling bias. Whereas Hopkins and Nunn's analysis compared parasite occurrences to known patterns of mammal richness, our analyses divided our parasite richness maps and point density into ten quantiles and took the difference between them to find hot- and cold-spots of parasite diversity based on a prediction from sampling point density (Figure 3.8). We also present a breakdown of our observed hotspots between species with and without human health relevance, showing the effect of sampling bias on parasite collections in Africa (Figure 3.9).

We discourage the unqualified interpretation of our results as an estimate of the underlying global patterns of parasite diversity. Compared to the 457 species of all parasites included in our study, there are an estimated 300,000 helminths alone, many to most of which have yet to be described by systematists. In future work, a reasonable assessment of parasite biodiversity hotspots might not be impossible, but it requires two major shifts in parasite open data. First, researchers must begin the process of georeferencing the major parasitology collections, including the full 70,000+ records of the U.S. National Parasite Collection (which we have georeferenced halfway in this study, and intend to complete by Fall 2017), and, more significantly, the 200,000+ records of the London Museum of Natural History (a more challenging task, as that analysis will require retrieving geospatial information from each published paper in the database hence why that analysis was not conducted during the timeframe of this study). Second, and just as importantly, targeted long-term work to profile the entire parasitic diversity of small, regional ecosystems is critically needed, especially in high-biodiversity systems that are proportionally neglected in our dataset. Included in that are highly-diverse regions like the Western Cape Province of South Africa, where parasitology work has been limited but is likely to discover high levels of uncataloged diversity ([218]); and conventional biodiversity hotspots like the Amazon basin that are hotspots of parasitological work and also assumed to be hotspots of parasite biodiversity, but where data is still too limited (see Figure 3.6). [219, 220] Finally, expanding all these analyses to freshwater systems and more significantly to the tremendous diversity of oceanic parasites (in particular for the speciose Cestoda, of elasmobranchs especially, and for unique specialists like Ozobranchus turtle leeches and cyamid whale lice that may be of special conservation interest) is a critical step forward.

Open data, model presentation and PEARL.

We present the final and forecasted distribution models for every species on an online server under the working title The Parasite Extinction Assessment and Red List version 0.1 (pearl.berkeley.edu; Figure 3.3). In the process of doing so, we set a new precedent for data storage in studies that do mass modeling of species range shifts, by making available a set of honest quantitative and qualitative metrics of data quality and model confidence:

- Data coverage: sample size by species is broken down into four quantiles and classified as "weak" (0-25%: 0-28 points), "fair" (25-50%: 29-42 points), "good" (50-75%; 43-80 points) and "excellent" (75-100%; 81-3289 points).
- Data uncertainty: the uncertainty radius (in meters) attached to each manually georeferenced point is available in the published form of the dataset, but many species include a combination of data from manually georeferenced sources (the USNPC and FeatherMites) and non-georeferenced sources (primarily GBIF). We set the uncertainty radius of non-georeferenced sources as zero, and then classified species' average uncertainty radius across zero and non-zero values into five categories: "perfect" (zero), "excellent" (0-5.1 km), "good" (5.1-6.5 km), "fair" (6.5-7.9 km) and "weak" (7.9-21.3 km).
- Model predictive accuracy: the true skill statistic (TSS; ranges from -1 to +1, where zero is a total lack of predictive power) and area under the receiver-operator characteristic curve (AUC; ranges from 0 to 1, with predictions greater than 0.5 considered better than random chance) for each final, weighted ensemble model.

While the last category is generally considered standard for publication-quality niche models, very few studies publish as many niche models simultaneously as this one, and presentation of results on online servers seldom includes these technical elements of model evaluation alongside the actual mapping efforts. With the quality of parasitological collections' geospatial information, this could not be more important to an honest and open scientific method. In niche modeling studies for poorly-documented clades like most parasitic groups, models can only be as good as the underlying data; rather than letting incomplete data preclude analysis (and thereby slow down parasite conservation and control efforts by at least 5-10 years), we instead recommend future work expanding our assessment's focus on similarly engaging in honest post-hoc evaluation of model accuracy and quality, and make such information available to the public. Doing so further helps highlight which species require the most thorough re-assessment in future incarnations of our tentative "Red List."

We also use our models to forecast the conservation status of each parasite in our study, using IUCN criteria to classify species into categories ranging from "Least Concern" to "Critically Endangered" based on average habitat loss rates. We adopt the breakdowns used by Thomas *et al.* [8] in terms of net habitat loss but drop the criteria involving absolute area, as suitable area for a parasite likely operates on very different dimensions than for a free-living

species, and their persistence via transmission is even more strongly linked to landscape connectivity. All our IUCN categorizations are also available at pearl.berkeley.edu. Expansion of PEARL beyond the 457 pilot species will follow the completion of the USNPC georeferencing project in late 2017, enabling higher-accuracy biodiversity mapping and parasite conservation planning in collaboration with other parasitology labs.

3.4 Results

We found that most parasites, like most free-living species [8, 139], face an existential threat from substantial habitat loss in a changing climate. Changes in habitat loss were affected slightly by differences in global climate models, but more pronouncedly by different climate scenarios, with a contrasting average native range loss of 20.2% in the most optimistic Representative Concentration Pathway scenario for greenhouse gases (RCP2.6) and 37.4% in the pessimistic (RCP8.5) scenario (see Figure 3.4). While no species lost its entire suitable range across every climate scenario, species lost an average of 29.0% of total habitat without dispersal, 86 species lost more than 50%, and eight lost more than 80% of their range. Even allowing for dispersal, 202 species still lost range by 2070, and 32 species lost more than half of their global suitable range; despite those losses, species gained an average of 16.2% suitable habitat, 29 species at least doubled the extent of their range, and seven at least tripled it (none of those seven having any evidence of zoonotic potential or human infection records).

Results highlight divergent outcomes of native range loss and range expansion for different species within the same clade (Figure 3.5). While previous literature has suggested that climate change may increase disease emergence through range expansion [152], we found that strictly-wildlife parasites experienced 17% more range gain, on average, than human infectious or biting species. This result could likely be explained by life history differences: in a two-way analysis of variance accounting for endoparasites vs. ectoparasites, the signal of human infection was insignificant, while endoparasites gained 39% more range than ectoparasites with dispersal ($p < 0.001$), and lost 10% less native range ($p < 0.001$). One potential explanation may be that the range limits of endoparasites are more commonly at disequilibrium due to dispersal limitations, while ectoparasites may already fill the majority of suitable habitat, relatively stabilizing their range size over time; similar assortative processes have proven important for shifting non-native plant species. [221]

Clade projections of total suitable range expansion with dispersal were significantly different ($F = 15.441$, $p < 0.001$), but no universal trends were visible across taxa. Lice had both the highest average native range loss and the highest average global range gain, although this conflicting result was likely a product of small sample size. Fleas and ticks consistently fared the worst, with both having average net loss even allowing for dispersal. The Thomas *et al.* [8] method for extinction rate estimation followed a similar pattern, with extinction rates for all species projected at 5.7%-9.2% without dispersal and 1.7-4.0% with it (see Table 2.1). A recent meta-analysis refined the global extinction risk estimate to 7.9% of free-living

species, placing our estimate of parasites' extinction rates in a comparable place to any other group. [139] Using a simplified version of the Thomas *et al.* habitat loss categories for IUCN red listing (which are, themselves, far reduced from actual IUCN listing criteria), 6.3% of species in our study are "endangered" and 0.7% "critically endangered" with dispersal, and 17% of species are endangered and 1.8% critically endangered without dispersal (Figure 3.10). Clade differences should be interpreted cautiously; our 457 species, while expansive for current data, is a meager subset of likely 300,000+ species of helminths alone [3]. An even more conservative analysis that only includes species with 50 or more points (rather than 20 or more) is available in the supplement (see Appendix 3). That analysis reduces the sample size from 457 species down to 196; the key results discussed above are essentially unaltered, although some specific extinction rates are predicted to be slightly higher (likely because extinction rate estimators converge on more conservative values as sample size increases).

In a 2008 study, Dobson *et al.* projected a coextinction rate of 3-5% for helminths in the next 50-100 years based only on host IUCN status [3], a far lower estimate than comparable projections for free-living vertebrates [8, 139]. A rough calculation using the same data shows that, if 15-37% of hosts were threatened (as Thomas *et al.* predicted), coextinction alone would be responsible for the loss of 8-24% of parasite species (Appendix 1) a far higher rate than has ever been observed empirically. The "paradox of missing coextinctions" is a significant problem on its own [125] , but compounded with the baseline extinction rates from habitat loss that we estimate here (see equation 1 in Appendix 1), we suggest a much higher fraction of species might be committed to primary and/or secondary extinction without dispersal by 2070: 5.6%15.4% of acanthocephalans, 11.9%29.0% of trematodes, 12.8%29.1% of cestodes, and 12.5%29.5% of nematodes (Figure 3.11). Therefore, the loss of parasite biodiversity could make a significant contribution to the sixth mass extinction, especially compared to the 7.9% baseline extinction rate suggested by Urban's recent meta-analysis [139].

3.5 Discussion

We adopt a more conservative methodology than Thomas *et al.* [8], likely producing lower estimated extinction rates. Our study used only species with 20 or more occurrences (thereby selecting disproportionately for cosmopolitan, generalist, and human-hosted parasite species with the largest ranges), ran maximum entropy models rather than using the BIOCLIM algorithm (the latter representing a method known to predict higher extinction rates compared to other niche modeling methods [188]), and the selected particular global climate models. Species with smaller ranges, which are disproportionately poorly sampled in our dataset, are likely subject to greater than average extinction risk. In addition, dispersal capacity will have a profound effect in determining which parasite ranges expand or contract, as some ectoparasites are likely to shift independent of host distributions [185] (leading to novel evolutionary opportunitiessee the Stockholm paradigm in evolutionary parasitology [222, 223]), but in other cases, hosts' shifting ranges may "escape" their parasites. [210,

211] Endoparasites with aquatic stages—like tapeworm coracidia, trematode cercariae or the drifting planktonic stages of copepods—or of long-distance dispersing hosts–like birds–may have greater dispersal capacity and ultimately fare better than average.

Range loss is also only one aspect of how parasites will experience climate change, and estimates of their vulnerability based only on range loss are likely to be fairly conservative. Even within a climatically suitable range, any given site may be missing hosts necessary for parasites to complete multi-stage life cycles; some parasites may be capable of plastic switching to truncated life cycles, but many will likely fail to persist in environments that ecological niche models would likely classify as "suitable habitat." [1] Even when host ranges shift in concert, phenological mismatch could prevent transmission from one host species to another, even for parasites that otherwise might appear to experience an overall gain in suitable range. [212] Finally, the transmission of parasites—and the interplay of virulence and host immunity—is often temperature-sensitive, and while it will have a critical role determining parasite vulnerability to extinction, it is also essentially impossible to predict using the types of models we present here. [1] In that way, parasite transmission ecology at a site-specific level could potentially produce even greater range losses than our models predict, once again making our models relatively conservative.

Even accounting for differences in sampling intensity between different continents, parasite species richness is far from evenly distributed at a global scale, potentially representing real underlying patterns or merely illustrating the sampling bias of parasite collections (e.g., substantially greater data completeness in North America). [224] Independent of potential biases, our results strongly support the hypothesis that climate change will drive a major redistribution of parasite biodiversity through habitat gains, losses, and shifts. Following a similar approach to previous work by Cumming *et al.*[150], we find that, in some cases, extinction is concentrated in the regions where our maps of parasite richness indicate diversity is the highest, as in the Gulf Coast of the United States, and in most of Western Europe (Figure 3.2). But at latitudes closer to the poles, where measured richness is lower, our simulations project species richness will double, triple, or increase even more. Many previous studies on Arctic parasite diversity have predicted and documented this developing ecological cascade [225, 226, 227, 228], but that phenomenon had yet to be predicted by broad-scale biogeographic models prior to this study. This result poses an alarming question: what will the consequences of a shifting wave of new parasite species, augmenting and possibly replacing native diversity, be for ecosystem stability, wildlife communities, and human health? Most parasites are not agents of emerging disease, but destabilized host-parasite networks might nevertheless create opportunities for new patterns of emergence. The redistribution of species ranges on a global scale is likely to create opportunities for the origin and evolution of new host-parasite pairings, as well as to change the regional balance of parasite diversity in different ecosystems; thereby allowing different parasitic taxa to become ecologically dominant, and potentially changing eco-evolutionary dynamics in the long term. [222]

While our study is the first to bring together this volume of data across parasite clades at the global scale, it is also a first step in a methodological progression that is ten years behind the cutting-edge of extinction analyses in nonparasitic animals. [229, 230] Current

best-practice research on climate change impacts on biodiversity depends on a cycle of data acquisition and model improvement that has progressed from basic biogeographic estimation methods to sophisticated biophysical mechanistic modeling that can encompass genetics, physiology, and dispersal. In the last decade, species-area relationship (SAR) models have been criticized as a method of assessing extinction risk, with strong evidence pointing to overestimation of extinction rates. [115, 118] Future analyses are critically needed to verify the stability of the species-area curve at continental scales for helminths and ectoparasites.

The possibility for parasites to experience compounded range loss with shifting host ranges, and their increased vulnerability relative to their hosts, is robustly independent of the species-area relationship. But parasite ecology desperately requires updated methods that can distinguish between likely winners and losers. The vast majority of climate-driven species extinctions may only proximally be a result of range loss, while population size, species interactions, presence of free-living life cycle stages, number of intermediate host species, and plastic and genetic components of climate tolerance and adaptation may be better predictors. [17, 231] Our study unambiguously deepens the need for experimental work, local long-term ecological experiments, and physiological mechanistic modeling to more accurately describe the threats parasites face than can be achieved with global distributional data.

Parasite conservation, as an applied discipline at the intersection of wildlife research and human health concerns, is in its infancy. Although parasite conservation is a topic of significant interest [5, 147] and has been for at least two decades [232, 233], the majority of parasitic biodiversity is unrecorded in ecological databases. Our study, together with the release of PEARL and the associated datasets, offers a foundational framework for including parasites in conservation ecoinformatics and biogeography. Threatened parasites require a specialized conservation approach, tailored to their unique life history, tremendous diversity, and the complex ecosystem services they provide. Some species in the most threatened clades may not even be parasites per se, such as the vane-dwelling feather mites in our study, for which the line between parasitism, commensalism, and mutualism is unresolved [234]; parasitic groups of nematodes are polyphyletic, and some of the nematodes in our dataset are similarly non-parasitic and free-living. [235] We include these species in our assessment nonetheless, because "parasitic clades" are dramatically understudied across the board, and their vulnerability was equally unassessed compared to their parasitic counterparts. [236] Similarly, protelean organisms, in which juvenile life stages are parasitic but adults are free-living (e.g., mites in the Parasitengona, or twisted-wing insects in the family Myrmecolacidae, in which the sexes are juvenile specialists separately on ants and orthopterans), are also likely to require extra conservation focus in a similar framework to ours.

Specialized, complex life cycles like these may face the most significant hidden vulnerability to habitat fragmentation or phenological mismatch [212], and may even experience a more severe extinction threat than our study predicts. While clade-level assessments may help identify some of these highly vulnerable groups, a finer-grained study of parasite extinction risk will require a more integrated perspective. Recent research has highlighted how parasites' range loss is likely to have a synergistic interaction with host specificity and host

functional traits (like thermal ecology and body size [1]). Combining the models presented here with Red List efforts for hosts, existing trait data, and network-based models of host parasite associations could substantially increase the resolution of parasite vulnerability assessments (including potential future iterations of PEARL). However, collecting the same big data that is revolutionizing wildlife and plant conservation will undoubtedly be a challenging next step for parasite conservationists.

3.6 Author Contributions

CJC, VMB, and KRB conceived the idea for the study; CJC, KRB, and CFC designed the study methods, and CJC, KRB, and AJP designed the data collection protocols. CJC, KRB, ERD, CFC, CAC, NCH, AJP, and VMB led data collection. TAD led collaboration with the London Natural History Museum; AJP led collaboration with the National Museum of Natural History, Smithsonian Institution. CJC and AJP led georeferencing and KRB, ERD, CFC, NCH, VMB, TD, and GC contributed to georeferenced data. JD, RJ, SM, and HP collected data for the feather mite project and GSC collected African tick data. CJC, KRB, CFC, ERD and OM contributed to model design; CJC ran all models and led data analysis. VMB compiled human infection data for parasites included in the final dataset. OM and CJC conceived the PEARL database, and OM and GC created the web infrastructure. All authors contributed to the writing of the manuscript and approved the final submission draft.

3.7 Acknowledgements

Our paper has benefited from mentorship, feedback and guidance concerning data collection, methods and analysis from a number of people, including (but not limited to) Sonia Altizer, Allison Barner, Jason Blackburn, Mike Boots, Robert K. Colwell, Andy Dobson, John Drake, John Harte, Kyrre Kausrud, Justin Kitzes, Britt Koskella, John Marshall, Timothe Poisot, Bree Rosenblum, Carl Schlichting, Dana Seidel, Patrick Stephens, Perry De Valpine, and an anonymous reviewer. We thank especially collaborators who offered to share data that were not presented in the final paper, in particular Kevin Lafferty and Giovanni Strona (FishPEST); Tim Littlewood and David Gibson (London Natural History Museum); Jean Mariaux (Musum D'histoire Naturelle, Geneva, Switzerland); Alejandro Francisco Oceguera Figueroa (Coleccin Nacional de Helmintos, Universidad Nacional Autnamo de Mxico); Sebastian Kvist (Royal Ontario Museum); Robert Poulin (unpublished data); Kevin Johnson (unpublished data); Jessica Light (unpublished data); Barry O'Connor (unpublished data); Agustin Jimenez-Ruiz (unpublished data); Mark Hafner (unpublished data); Kurt Galbreath (unpublished data) and Kayce Bell (unpublished data). Special thanks goes out to Kathryn Ahlfeld and William E. Moser at the National Museum of Natural History for assistance in systematics work on the USNPC. We also thank the undergraduates at UC Berkeley and the University of Connecticut who have contributed to this project over the

past two years, including Faith De Amaral, Humza Siddiqui, Fred Heath, Savannah Miller, and Nicole Kula; and Smithsonian intern Sarah Fourby. We finally thank every researcher who contributed specimens to the museum collections that made this analysis possible. This project was funded in part by the UC Berkeley Department of Environmental Science, Policy and Management and by the A. Starker Leopold Chair held by WMG, and by project CGL2015-69650-P and Ramon y Cajal research contract (RYC-2009-03967) to RJ.

3.8 Appendix 1. Primary, Secondary and Compounded Extinction Rates

Dobson *et al.* [3] provide the following values for mean host specificity (top) and parasite species richness (bottom):

	Chondrichthyes	Osteichthyes	Amphibia	Reptilia	Aves	Mammalia
Trematoda	2	6.35	5.4	1.77	2.97	2.01
	(51)	(5,831)	(1,170)	(3,773)	(9,862)	(3,714)
Cestoda	1.69	6.38	4.75	2.21	2.36	1.89
	(1,352)	(4,466)	(283)	(1,112)	(14,058)	(4,637)
Acanthocephala	—	14.95	6.74	12.5	8.35	4.32
	—	(1,226)	(140)	(212)	(779)	(301)
Nematoda	2.67	10.28	5.27	2.12	3.28	6.07
	(152)	(2,631)	(2,662)	(6,389)	(9,150)	(2,979)

Using a formula for coextinction rates, some simple math allows an updated estimation from different levels of host extinction risk, based on other estimates than IUCN data. Koh *et al.*'s [122] method for affiliate extinction probability from host risk E and specificity s estimates:

$$\bar{A} = (0.35E - 0.43)E \ln S + E \tag{3.5}$$

Plugging in a 15-37% extinction risk for hosts (an extreme scenario) and the values for host specificity gives a best case scenario:

	Chondrichthyes	Osteichthyes	Amphibia	Reptilia	Aves	Mammalia
Trematoda	11.08%	4.53%	5.45%	11.77%	8.84%	11.05%
Cestoda	12.03%	4.51%	6.18%	10.51%	10.14%	11.40%
Acanthocephala	—	0%	4.20%	0.70%	2.98%	6.71%
Nematoda	9.44%	1.81%	5.59%	10.75%	8.27%	4.79%

And a worst case scenario:

	Chondrichthyes	Osteichthyes	Amphibia	Reptilia	Aves	Mammalia
Trematoda	29.29%	16.45%	18.25%	30.65%	24.90%	29.24%
Cestoda	31.17%	16.40%	19.68%	28.18%	27.45%	29.92%
Acanthocephala	—	6.93%	15.79%	8.92%	13.40%	20.73%
Nematoda	26.08%	11.09%	18.52%	28.65%	23.79%	16.95%

Weighting these each by the relative richness of different host groups gives the final assessment:

Clade	Richness	CI based on Thomas *et al.*
Trematoda	24,401	8.4% to 24.1%
Cestoda	25,098	9.5% to 26.1%
Acanthocephala	2,658	1.9% to 11.0%
Nematoda	23,963	7.5% to 22.3%

And a total: 76,930 species of helminths with a weighed estimate of 8.3% to 23.8% extinction rate. If specialist parasites are disproportionally hosted by low-risk species, this may explain some of the reason parasite extinction is less prevalent than predicted. [3] But a more parsimonious explanation is that the low projected rate comes from the use of incomplete IUCN red list data that underestimates host vulnerability.

We focus in the above analysis on providing a less conservative estimate of how extinction rates might compound with host vulnerability, and maintaining a consistent estimate based on the Thomas et al. SAR for both hosts and parasites. However, we could just as easily implement the same analysis using the conservative meta-analysis based figure Urban recently published [139]. Estimates based on the SAR suggest a 22% extinction rate and estimates based on expert opinion are similarly high. But Urban suggests an overall extinction baseline across plants and animals of approximately 7.9%. Using that metric, we can once again calculate by-group estimates, but we can refine it even further by using Urbans by-group estimates for amphibians (12.9%), reptiles (9.0%), mammals (8.6%), fish (7.6%, using the same value for chondrichthyes and osteichthyes) and birds (6.3%). We can even use the 95% credible interval, once again, to generate a best-case scenario table:

	Chondrichthyes	Osteichthyes	Amphibia	Reptilia	Aves	Mammalia
Trematoda	2.41%	0.77%	2.61%	4.21%	2.47%	4.21%
Cestoda	2.65%	0.77%	3.03%	3.71%	2.90%	4.36%
Acanthocephala	—	0%	1.89%	0%	0.54%	2.37%
Nematoda	2.00%	0.09%	2.69%	3.80%	2.29%	1.54%

And, a worst-case scenario:

	Chondrichthyes	Osteichthyes	Amphibia	Reptilia	Aves	Mammalia
Trematoda	11.93%	4.98%	7.76%	10.71%	5.16%	9.59%
Cestoda	12.94%	4.95%	8.67%	9.55%	5.99%	9.90%
Acanthocephala	—	0%	6.17%	0.48%	1.41%	5.74%
Nematoda	10.19%	2.08%	7.93%	9.77%	4.80%	4.02%

And, aggregated by group:

Clade	Richness	CI based on Urban *et al.*
Trematoda	24,401	2.61% to 6.79%
Cestoda	25,098	2.82% to 7.05%
Acanthocephala	2,658	0.31% to 1.35%
Nematoda	23,963	2.40% to 6.11%

These numbers may represent a more literature-based estimation of helminth co-extinction rates, but we present the Thomas *et al.* based numbers in the main text to maintain consistency of methods across hosts and parasites, and to present a true worst-case scenario for how severe the threats parasites face might become.

To determine how these different projections affect total extinction rate projections, we implement a combinatorics formula that assumes (with no prior knowledge) zero covariance between host and parasite extinction from climate change (or, more accurately, no covariance between primary and secondary extinctions):

$$P(extinction) = 100\% - (100\% - p_{directextinctionfromclimatechange})(100\% - p_{coextinction}) \quad (3.6)$$

Giving us the estimates presented in the main text with dispersal (in percentages):

	1° Extinction	2° (Urban)	1° + 2° (U)	2° (Thomas)	1° + 2° (T)
Trematoda	(0.11, 1.2)	(2.61, 6.79)	(2.72, 7.91)	(8.4, 24.1)	(8.50, 25.0)
Cestoda	(0.07, 0.07)	(2.82, 7.05)	(2.89, 7.12)	(9.5, 26.1)	(9.56, 26.15)
Acanthocephala	(0.21, 0.60)	(0.31, 1.35)	(0.52, 1.94)	(1.9, 11.0)	(2.11, 11.53)
Nematoda	(1.3, 3.3)	(2.40, 6.11)	(3.67, 9.21)	(7.5, 22.3)	(8.70, 24.86)

And, without dispersal:

	1° Extinction	2° (Urban)	1° + 2° (U)	2° (Thomas)	1° + 2° (T)
Trematoda	(3.8, 6.0)	(2.61, 6.79)	(6.31, 12.38)	(8.4, 24.1)	(11.88, 28.65)
Cestoda	(3.6, 4.0)	(2.82, 7.05)	(6.32, 10.77)	(9.5, 26.1)	(12.76, 29.06)
Acanthocephala	(3.8, 4.9)	(0.31, 1.35)	(4.10, 6.18)	(1.9, 11.0)	(5.63, 15.36)
Nematoda	(5.4, 9.3)	(2.40, 6.11)	(7.67, 14.84)	(7.5, 22.3)	(12.50, 29.53)

We note one particular difference of interest to researchers by incorporating the covariance between different host group extinction rates and their specificity, the gap between acanthocephalans and other more threatened helminths emerges very clearly.

Future work relating coextinction to parasite primary extinction rates will require an approach that links host and parasite distributions and accounts for the missing covariance. While a few odd winners in any group will have the pre-existing niche breadth to benefit from climate change, theory predicts that the majority of species should suffer at least partial range loss both hosts and parasites included and at its most extreme this means that vulnerability should compound across parasites and obligate hosts. In our study, fragmented host information and the heavy bias towards agricultural and human-infectious species makes such an approach impractical or uninformative; however, we outline a targeted approach for subsequent studies that focus on smaller specialist clades. Key to that approach is simulating the simultaneous shift of hosts and parasites and searching for potential discrepancies between their ranges, an approach notably used by Pickles *et al.* [212] Host information is readily available for mammals and many reptiles and amphibians from the IUCN, and for birds by BirdLife International range maps. Projecting the joint shift of hosts and parasites can be used to calculate a host-constrained projection of parasite future ranges, which accounts for potential independence in shifting habitat suitability.

This approach also allows another, more conceptual approach to exploring parasite vulnerability; in this approach, the Thomas *et al.* method can be implemented for hosts alone (following an identical procedure for projecting range shifts) and converted into parasite vulnerability using Koh *et al.*'s method for affiliate extinction probability from host risk E and specificity s (which can be calculated from host-parasite association network data). This host-as-proxy regional estimation can then be compared against a reimplementation of the Thomas *et al.* method using parasite areal changes (with or without host ranges as constraints, corresponding to total and intrinsic vulnerability to extinction). We term the relative fraction of vulnerability driven by hosts (the Koh-converted extinction risk divided by the constrained parasite-based extinction risk) the compounded risk factor and suggest that future analyses using our global parasite database could explore how much greater than 1.0 those values are for parasites with different levels of specificity in different ecosystems. For highly specialized species with a single host, the extinction rates should have a rough 1:1 correspondence; however, generalist and parasitic species with several free-living stages should be more severely affected by their intrinsic vulnerability to extinction and have a smaller proportional compounded risk factor.

3.9 Appendix 2. Canonical and Maximum Entropy Approaches to the Species Area Relationship

In the canonical Arrhenius species area relationship (SAR), $S = cA^z$,a slope z of 0.25 provides a convenient solution to extinction rate estimation in the Thomas *et al.* method. Despite the substantial criticism the method has faced, especially in a seminal series of papers by Harte and Kitzes [118, 114, 113], the canonical SAR is still valid under a very specific set of circumstances. Harte and Kitzes suggest that for an island SAR in particular, the slope

can be predicted as 0.25 if "(i) total abundance in the new area A is proportional to area, (ii), individuals found in A are chosen by a random draw of all individuals in A_0, and (iii) the number of individuals of each species in A_0 follows a canonical lognormal abundance distribution." [113] Moreover, they draw a parallel to predicting extinction from climate change, recommending the nested SAR is more applicable to loss of native range (shrinking suitable habitat of an entire region) while the island SAR may be more applicable to ranges shifting into novel habitat.

There is, in the Harte & Kitzes method (the Maximum Entropy Theory of Ecology, or METE), a method for deriving a nested SAR that accurately predicts its curvature towards z = 0 at continental scales. They define a probability of survival P (compared to the original probability of survival P0) that, in the original canonical SAR, is

$$P = P_0(A/A_0)^z \tag{3.7}$$

The METE formulation of P accounts for initial abundance n0 and suggests species face certain extinction when the ratio of abundance n and n0 drops below a threshold r_c (i.e., the pseudo-extinction threshold or minimum viable population). For a single species,

$$P\big(\frac{n}{n_0} \leq r_c\big) = 1 - \frac{\big[\frac{n_0\beta}{1+n_0\beta}\big]^{r_c n_0} - \big[\frac{n_0\beta}{1+n_0\beta}\big]^{n_0}}{(1+n_0\beta)\ln(1+\frac{1}{n_0\beta})} \tag{3.8}$$

where they substitute β for A/A_0. This approach allows direct calculation of an extinction area relationship by evaluating each species fate (or, by making top-level assumptions about abundance distributions in the community). We observe that analyses with uncertainty about abundance distributions and viable population sizes could easily be implemented in a Bayesian framework, with basic priors assumed for the demographic free parameters.

To do that, or to use the METE approach in general, requires assumptions about the population trajectory and aggregation of parasites, the distribution of their abundances within a community (log-normal or not), and the critical population size below which extinction is certain. For endoparasite helminths, critical population sizes might be easily solved through conventional epidemiological methods. The host density threshold is a frequently used metric in epidemiology, and basic assumptions about parasite aggregation within a single infected individual might make relating the HDT to r_c readily possible. But the data to inform such assumptions is absent at broad scales in parasite ecology; only a couple or a few species in our study have such data.

In summary, implementing the SAR to predict extinction for parasites is unprecedented, and so poses a number of problems. The applicability of the SAR with a slope of 0.25 or higher to parasites is assumed (given its applicability to their hosts), and is supported for use in our study by the limited literature applying SARs to parasites. [237, 238] We make the explicit choice to adhere to the Thomas *et al.* implementation of the SAR approach with $z = 0.25$ to avoid further entangling our estimates in unsubstantiated assumptions about parasite demography, or about how parasite aggregation among hosts (which can

follow a negative binomial distribution, and in the context of climate change, will be non-independent from host area and abundance declines) would affect the validity of the METE. The derivation and empirical validity of the METE has already been the subject of one book and numerous articles, and exploring its applicability to parasites could require similar multi-year efforts, using data that is by-and-large missing from current parasitology databases. Harte and Kitzes suggest three major tasks to refine their methods:

1. "Develop better projection methods for the number of species shared among sets of disjointed habitat patches."

2. "Enrich understanding of the shapes and slopes of SARs at large spatial scales."

3. "Enrich understanding of secondary species losses due to trophic web-induced and other interaction-induced cascades."

and we concur that these are critical tasks before the SAR and extinction area relationship for parasites can be better refined beyond the Thomas *et al.* methodology.

3.10 Appendix 3. A More Restrictive Analysis, Based on 50+ Point-per-Species Models

Sample size is an important limiting factor in all ecological niche modeling (ENM), and detailed attention to the role sample size plays in model accuracy is a key part of due diligence for researchers building and applying ENMs. In our analyses in the main text, we present results for species that have a minimum of 20 unique occurrences. However, in other work, 50 or more occurrences is a more stringent threshold that some might use. Here, we present the key analyses from our main text, re-analyzed for the subset of species with 50 unique occurrences. That reduces the sample size from 457 species down to 196; at the clade scale, the effects of that reduction are most apparent.

In the restrictive analysis (versus the 20 point analysis in the main text), climate scenarios have essentially the same effect on habitat loss, with an average native range loss of 21.4% (vs. 20.2%) in the optimistic RCP 2.6 scenario, and of 41.2% (vs. 37.4%) in the pessimistic RCP 8.5 scenario. Across scenarios, species lose an average of 31.5% (vs. 29%) of total habitat without dispersal. Of 196 species, 36 lost more than 50% of their range, and one lost more than 80% of its range. Even allowing for dispersal, 106 of 196 (versus 202 of 457) species lost range by 2070, and 14 species lost more than half of their global suitable range; despite those losses, species gained an average of 0.3% suitable habitat (vs. 16.2%; the only noticeable difference from the main analysis); four species doubled the extent of their range, and none tripled. Strictly-wildlife parasites experienced an average of 10% more range gain (vs. 17%) than zoonoses. That effect still originates in endoparasite vs. ectoparasite differences, with endoparasites gaining 31% (vs. 36%) more range than ectoparasites with

dispersal ($p < 0.001$), and losing 12% less native range (vs. 10%; $p < 0001$). Clade differences are still significant (one-way ANOVA: $F = 8.287$, $p < 0.001$ vs. $F = 15.441$, $p < 0.001$).

Recreating Table 2.1 for this analysis yields roughly comparable results (Table 2.2). The most significant difference between these analyses comes for mites and lice, which experience substantially less gain, and for which the sign of average habitat loss is flipped from positive to negative (and the upper bounds of potential habitat gain are substantially reduced). This is likely due to the reduction of sample size to 2 species each for both clades, on which grounds, we note that the results marked with an asterisk should likely be entirely disregarded (especially as the extinction estimators are unlikely to be at all meaningful). More generally, restricting sample size even further produces a minor increase in habitat loss, and therefore extinction rates; overall, the patterns of extinction risk are comparable, with 7.2-9.8% of species committed to extinction without dispersal (vs. 5.7%-9.2%), and 2.3%-4.6% with dispersal (vs. 1.7%-4.0%). For our "IUCN classification" analysis, with dispersal, none would be critically endangered (vs. 0.7%), 7.1% (vs. 6.3%) of species would be endangered, 26.5% (vs. 18.8%) vulnerable, and 66.3% (vs. 74%) least concern; without dispersal, 0.5% (vs. 1.8%) would be critically endangered, 17.8% (vs. 17.1%) endangered, 42.8% (vs. 49.5%) vulnerable, and 38.8% (vs. 31.7%) least concern, continuing to reflect an overall subtle increase in risk associated with the restricted sample size.

For poorer-sampled clades (Astigmata and Phthiraptera, and to a lesser extent, Acanthocephala and Cestoda), reducing the sample size is likely to have substantially reduced the validity of the analyses we present in the main text. More generally, it may be the case that models with more occurrence data better capture the equilibrium realized niche of the species, and therefore find less novel habitat for species to expand into. Alternatively, it may be that other subtle biases in data collection (such as spatial autocorrelation between sample sites) have produced more detailed data for species with overall more restricted niches or ranges. Speculation as to the mechanisms of the pattern is likely to be unsuccessful given the combination of data sources assembled in the study, each contributing their own intrinsic pattern of sampling bias. However, the more restricted analysis only further confirms that every group in our study is likely to have a handful of species experiencing devastating range loss, leading to significant extinction risk.

As a final precaution, we compared accuracy metrics for models under and above 50 points, to examine whether models might be failing using the 20-or-more criterion. A very small effect is detectable in the AUC (mean under 50: 0.945; mean over 50; 0.948; $t = -3.235$, $df = 8221.5$, $p = 0.00122$, $95\% CI = (-0.0051, -0.0013)$), but AUC is also a comparatively unreliable metric of model performance, and the overall effect is minimal as both groups appear to perform extremely well on average. (Moreover, the lowest AUC recorded in the entire study is 0.728, which is still within the range of well-performing, published models, and certainly gives no indication of objective model failure.) An opposite effect is detectable for the true skill statistic (means 0.815 vs. 0.797; $t = 6.6715$; $df = 08053.5$; $p < 0.0001$; $95\% CI = (0.0127, 0.0232)$), for which models under 50 points appear to perform slightly better, but again, with minimal effect. Based on both of these simple tests, we find no strong evidence to suggest models with between 20 and 50 points perform noticeably poorly

and might introduce non-trivial error into our main results (again, noting that 20 or more points is a common threshold in the literature; see also [169]).

Given that extinction estimates for at least two clades (Astigmata and Phthiraptera) would become invalid with a more restricted analysis, and others would likely have been noticeably weakened, we elected to present results drawn from all 457 species in the main text. Our study highlights the challenges of data availability in parasitology research; the fact that two clades extinction rates would be essentially unmeasurable with a more restrictive sample size rule only highlights those challenges. However, we note that researchers interested in using the results of the more restricted analysis can find them here, and can also obtain individual sample sizes and accuracy metrics for every species and model in the supplemental datasets.

Figure 3.1: Final dataset breakdown by source and clade.

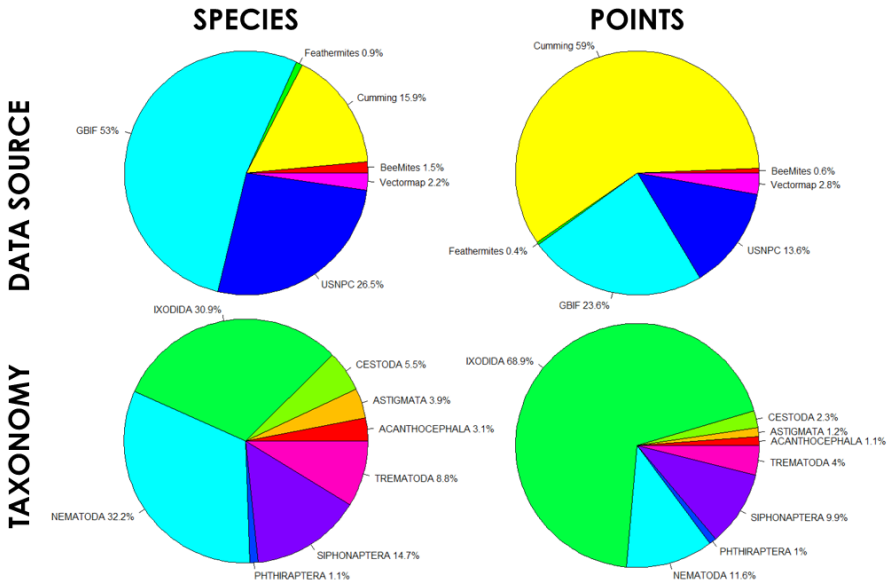


Figure 3.2: Gradients of species richness and predicted turnover through extinction and redistribution. a. Current distribution of parasite species richness (S) in our dataset is calculated by stacking binary outputs of species distribution models (see point distributions in Figure 3.6). b. Turnover (in species units) is measured by following the same procedure from 18 combinations of GCMs and RCPs for year 2070, and taking the average difference (ΔS) from 2016. c. Proportional change (ΔS/S) is most severe in low-diversity areas where parasite richness is predicted to increase as a consequence of latitudinal shifts.

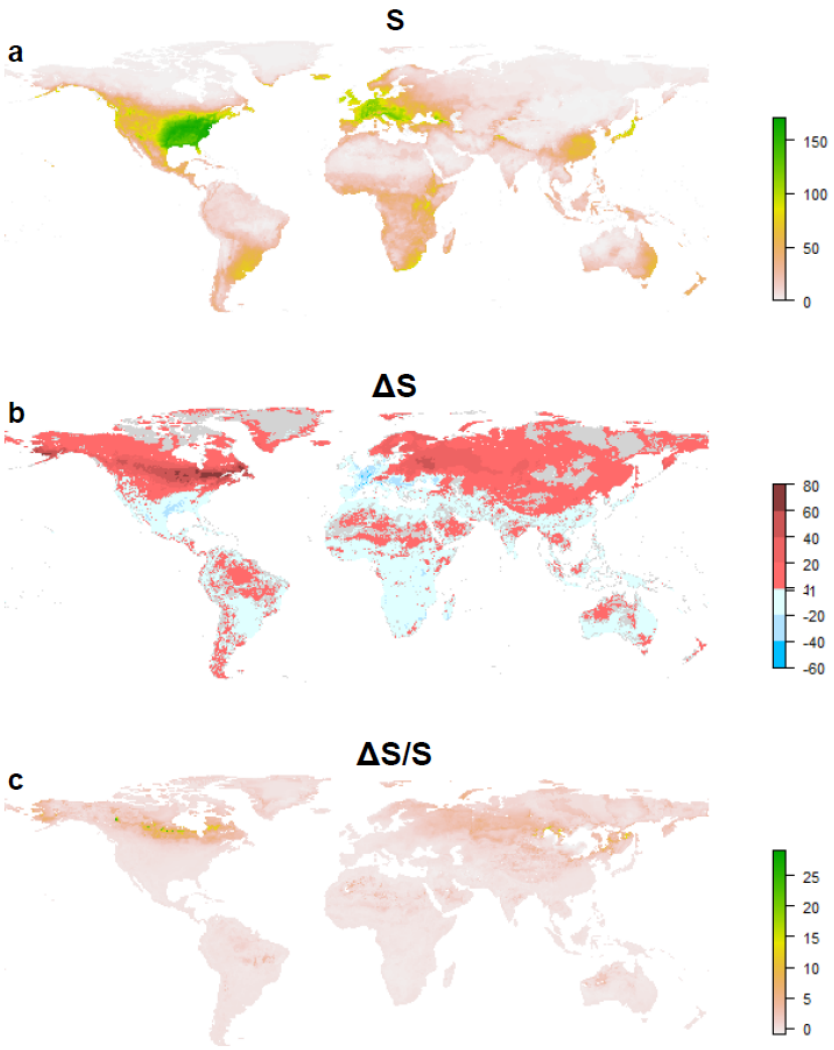


Figure 3.3: Example presentation of species distribution and conservation status on the Parasite Extinction Assessment & Red List. Real results are shown for *Abbreviata bancrofti* (Nematoda), a representative species in our study and the first available on the website alphabetically.

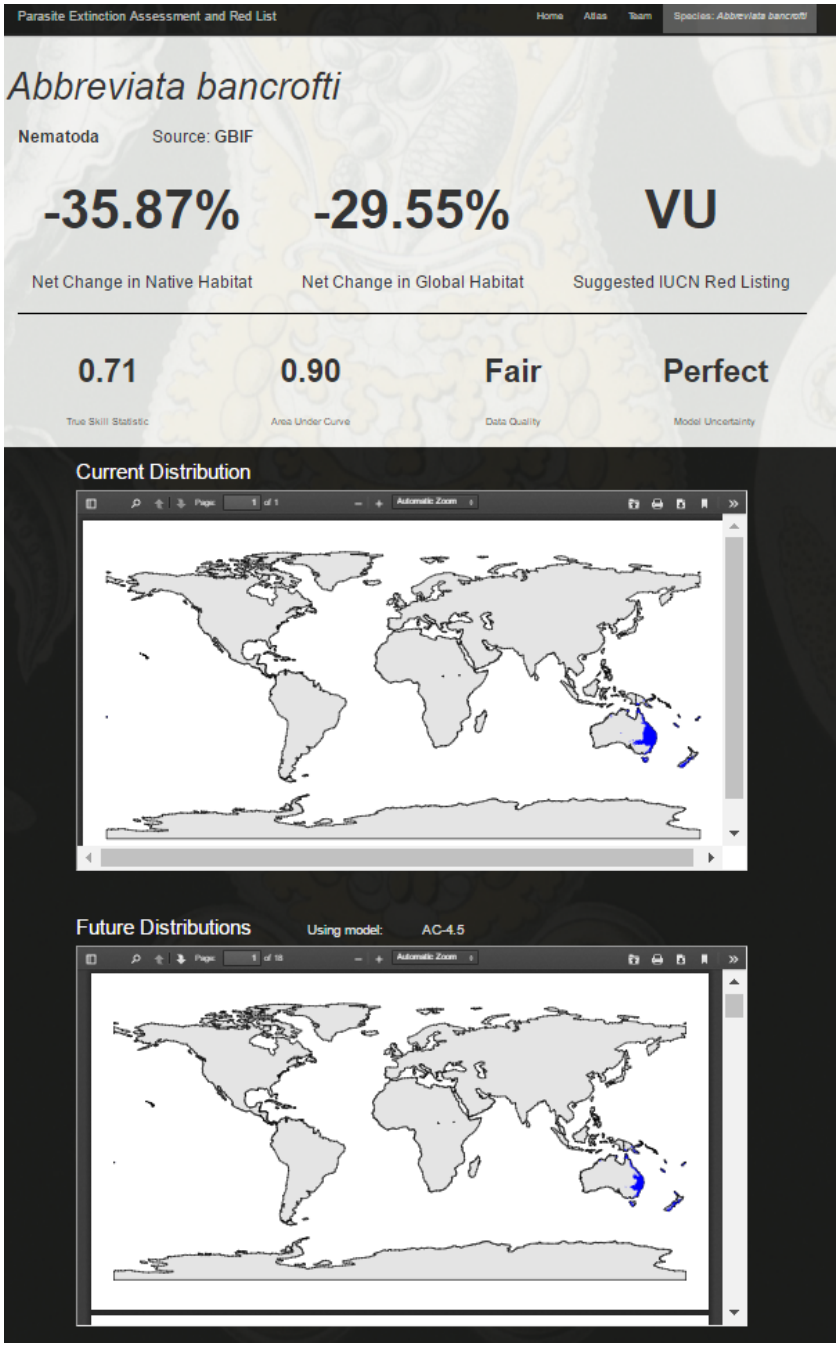


Figure 3.4: Loss of native habitat broken down by RCP and GCM. Results are broken down into all models and the subset of models that "perform well" (with a true skill statistic over 0.6).

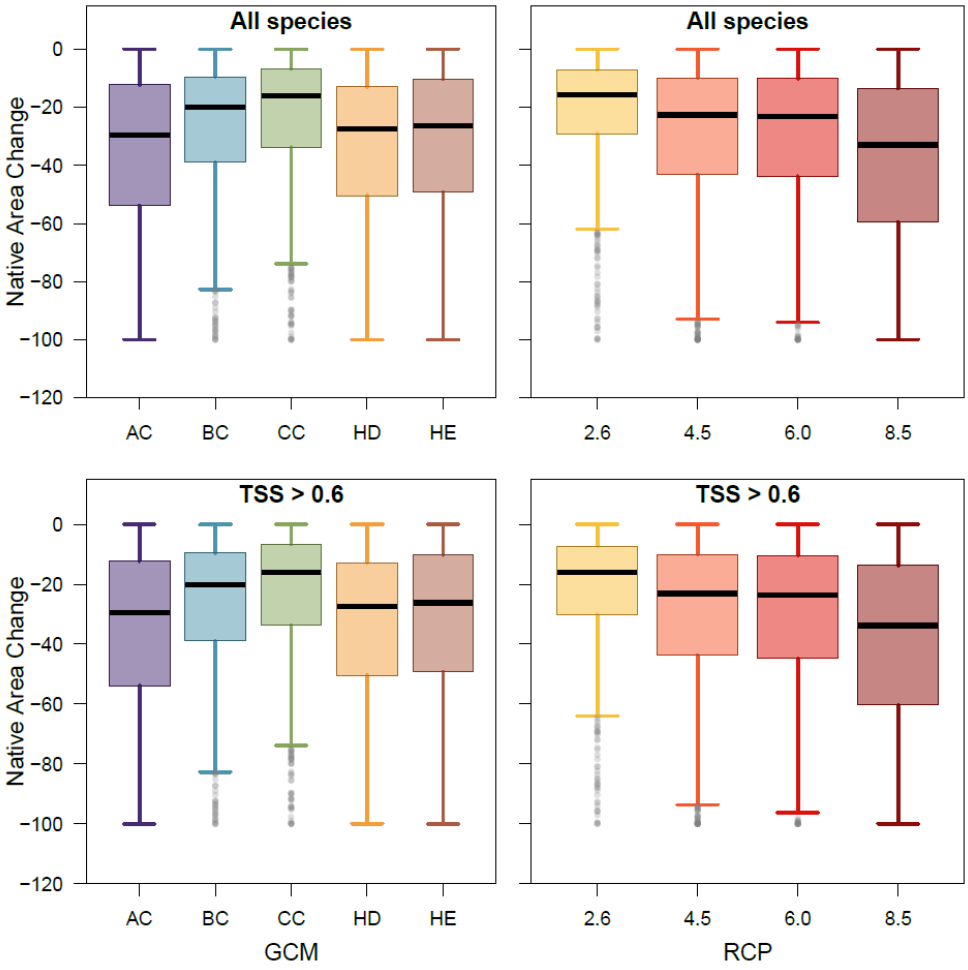


Figure 3.5: Tradeoffs between biodiversity loss and emergence across parasite clades. Discrepancies between current and future range size are projected as averages across all GCMs and RCPs at the species level, with (y-axis) and without (x-axis) dispersal, and broken down by our eight clades. Most clades are likely to be subject to moderate-to-extreme range loss; but the species with projected extreme expansions are mostly helminth endoparasites (in particular, nematodes and trematodes).

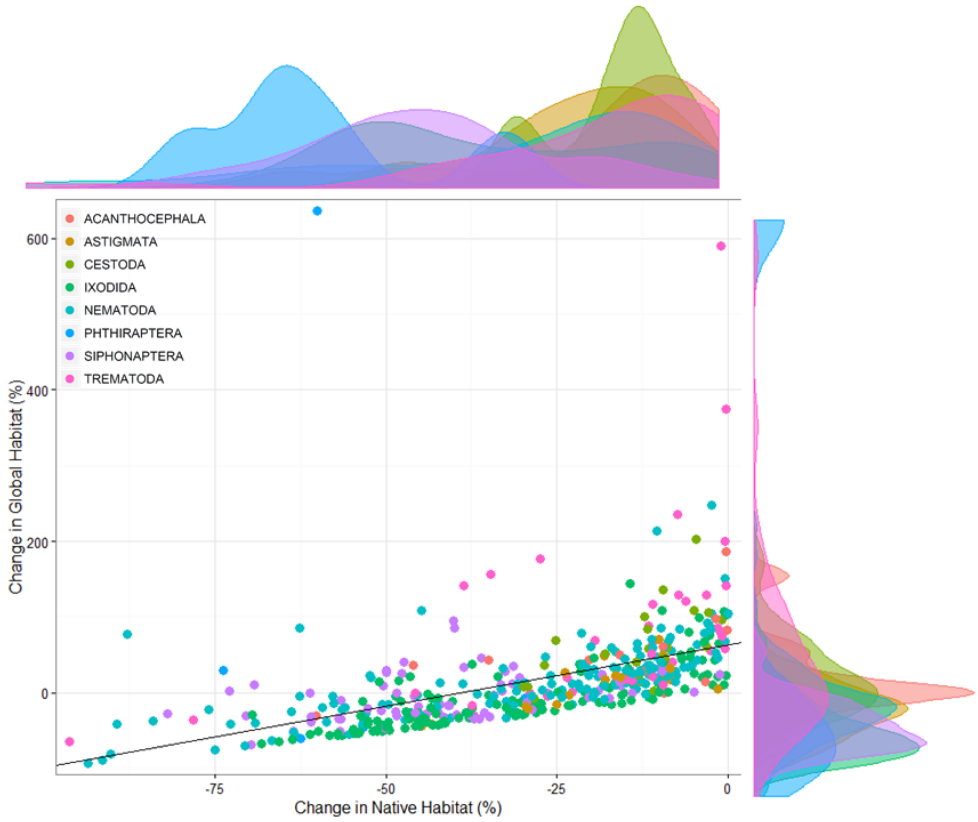


Figure 3.6: Sources and distribution of occurrence data. a. Data from the US National Parasite Collection (grey: not included in final dataset; black: included in the study based on minimum sample size, taxonomic cleaning, etc.). b. Data from VectorMap (blue) and the Global Biodiversity Informatics Facility (orange). c. Data from the Bee Mites database (blue), the Cumming tick database (red), and georeferenced data of the feather mite database (black).

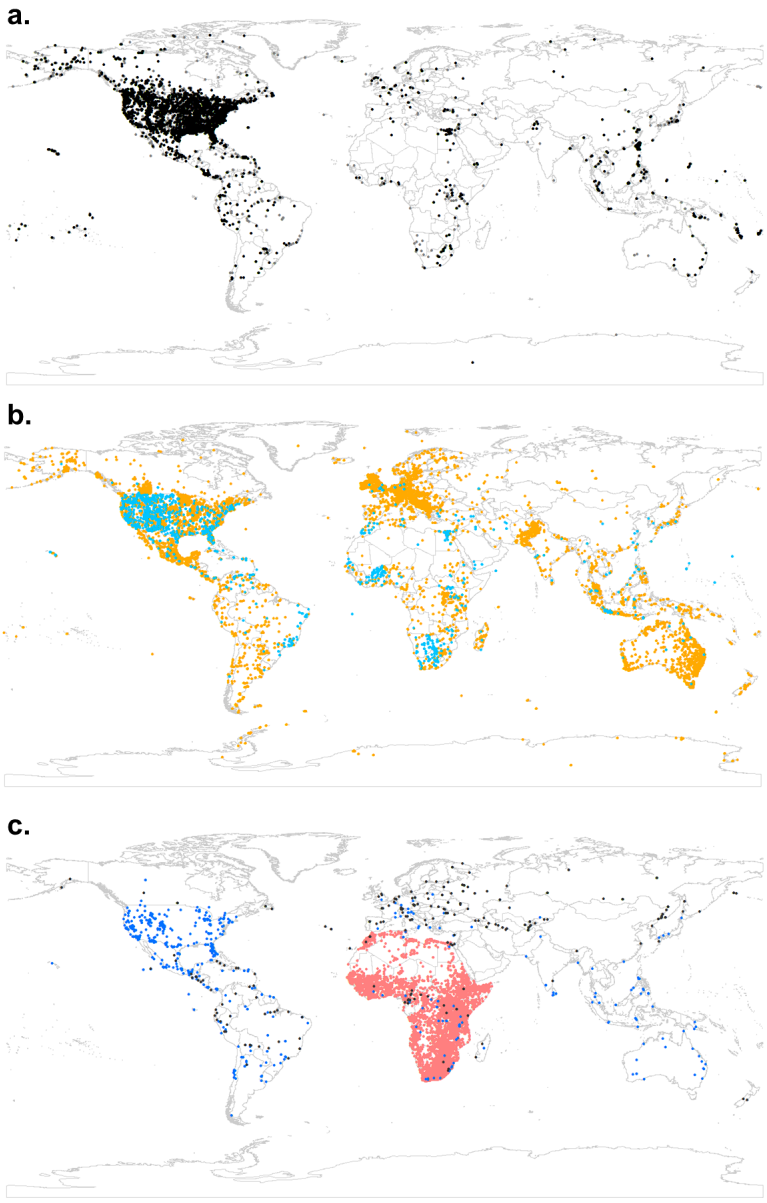


Figure 3.7: Loss of native habitat broken down by feature classes and regularization multiplier. Results are broken down into all models and the subset of models that "perform well" (with a true skill statistic over 0.6). Models are built from a combination of five feature classes: linear (L), quadratic (Q), hinge (H), product (P), and threshold (T).

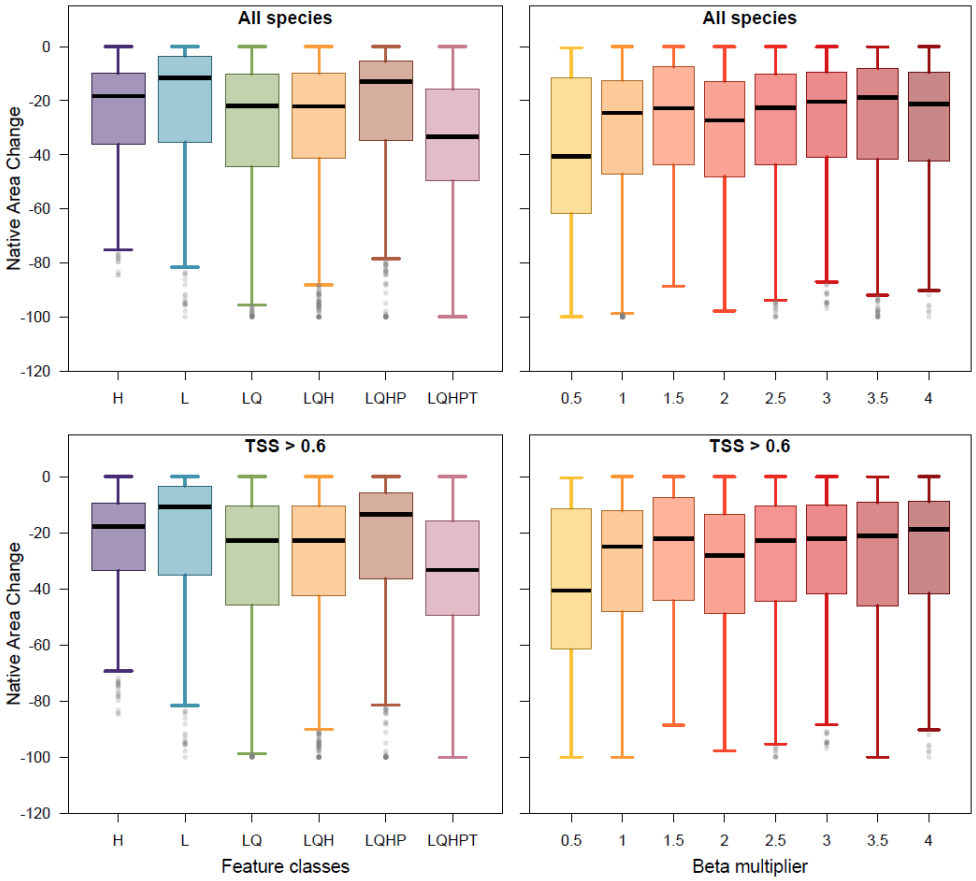


Figure 3.8: Visualizing spatial bias in species richness gradients. a-c, From the distribution of points included in our global parasite database, we constructed a global compiled map of species richness (a) calculated by layering every species distribution model. But with biased sampling that map may reflect false patterns; so we also present the density of points smoothed with a Gaussian filter with $\sigma = 1$ (b), and subtract the latter from the former to show richness relative to sampling intensity (c).

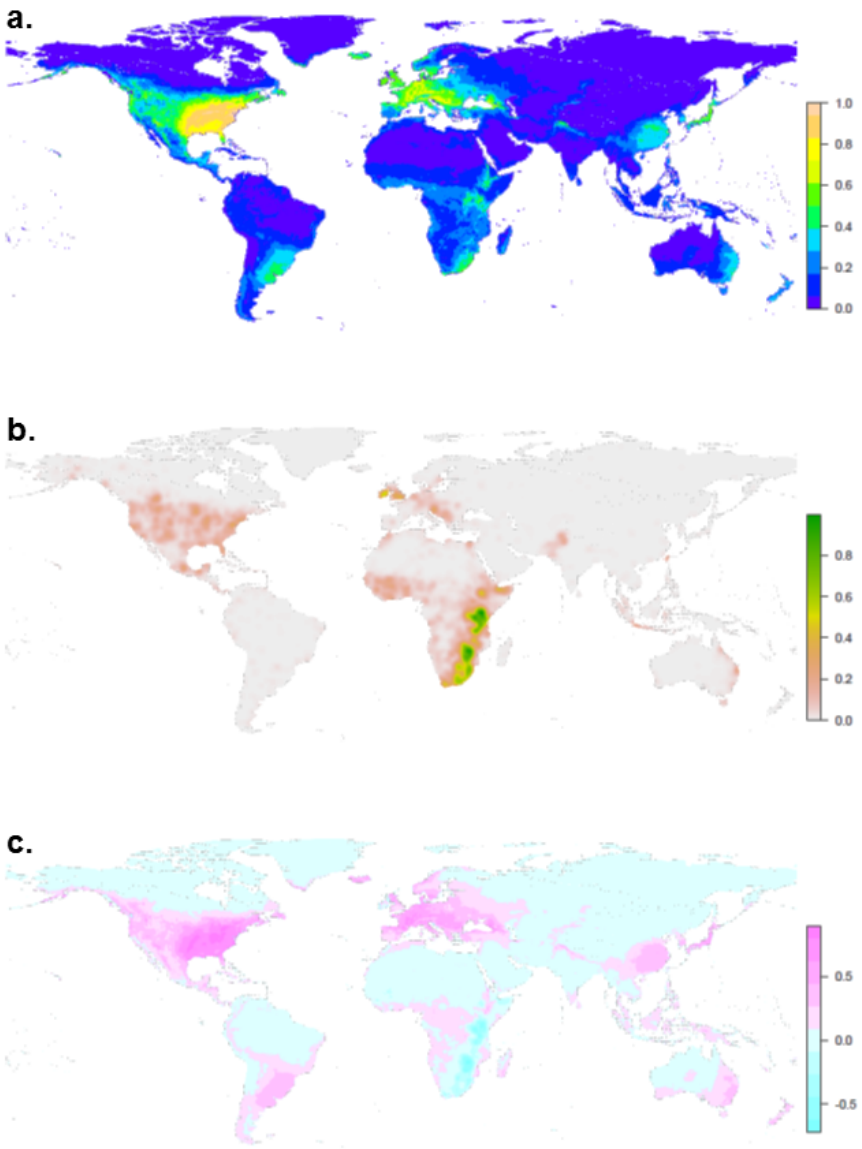


Figure 3.9: Parasite richness gradients by human health concern. a, Species richness gradients for species in our study with human health relevance (zoonotic endoparasites and ectoparasites with records of feeding on humans) compared to b, richness gradients for strictly-wildlife or free-living species.

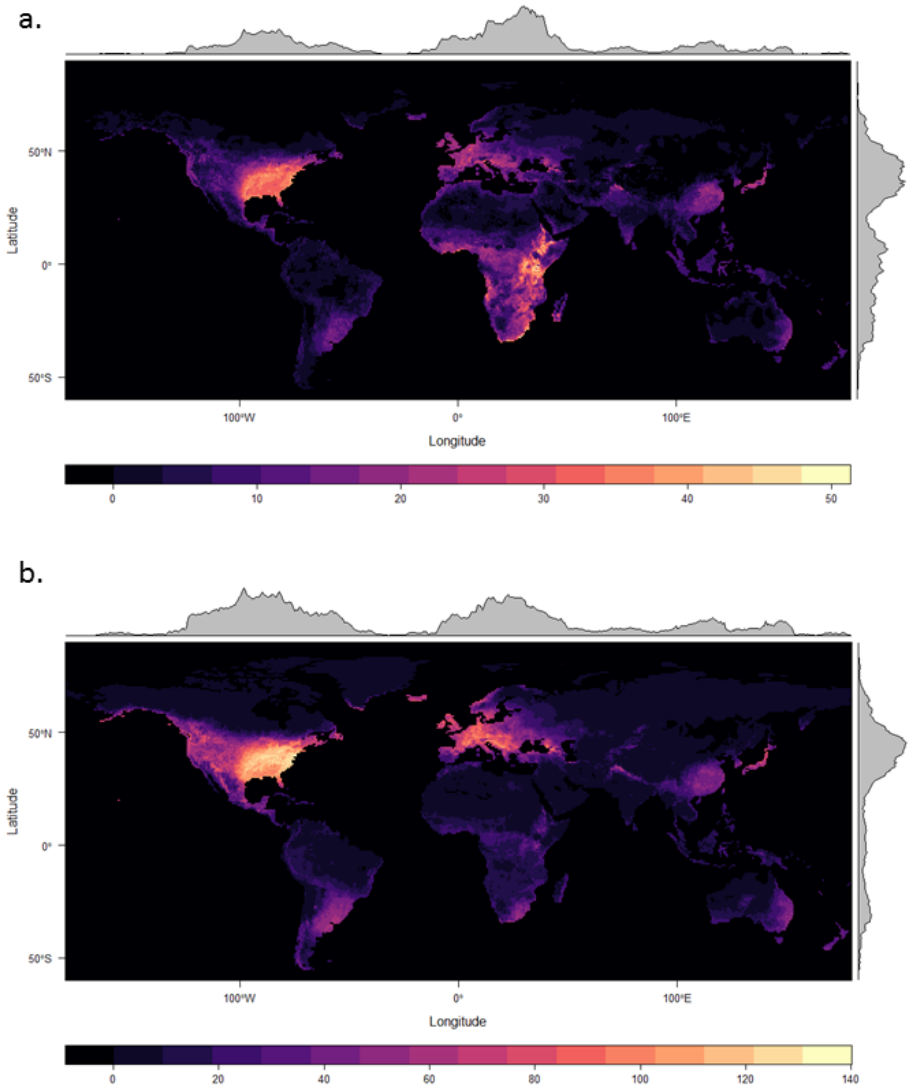


Figure 3.10: Comparative IUCN "Red List" breakdowns by clade. a. Breakdowns are given by habitat loss categories from now to 2070: 0-25%: least concern; 25-50%: vulnerable; 50-80%: endangered; 80-100%: critically endangered. b-i. Conservation classifiers are broken down for eight major clades: b. Acanthocephala (n = 14 spp.); c. Astigmata (n = 18); d. Cestoda (n = 25); e. Ixodida (n = 141); f. Nematoda (n = 147); g. Phthiraptera (n = 5); h. Siphonaptera (n = 67); i. Trematoda (n = 40).

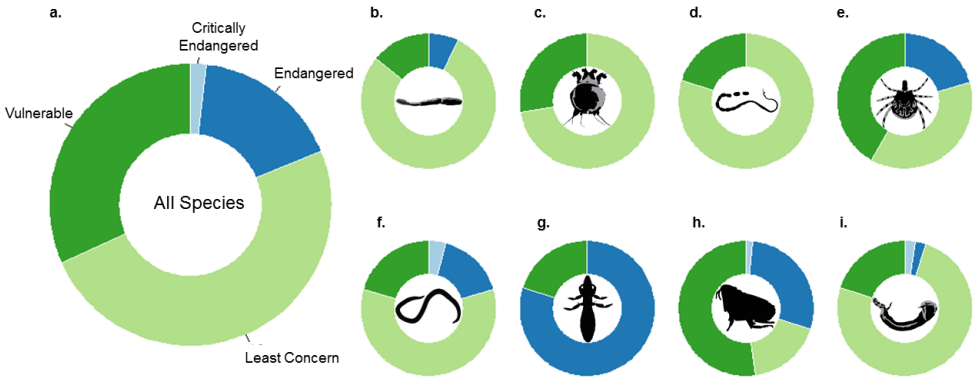
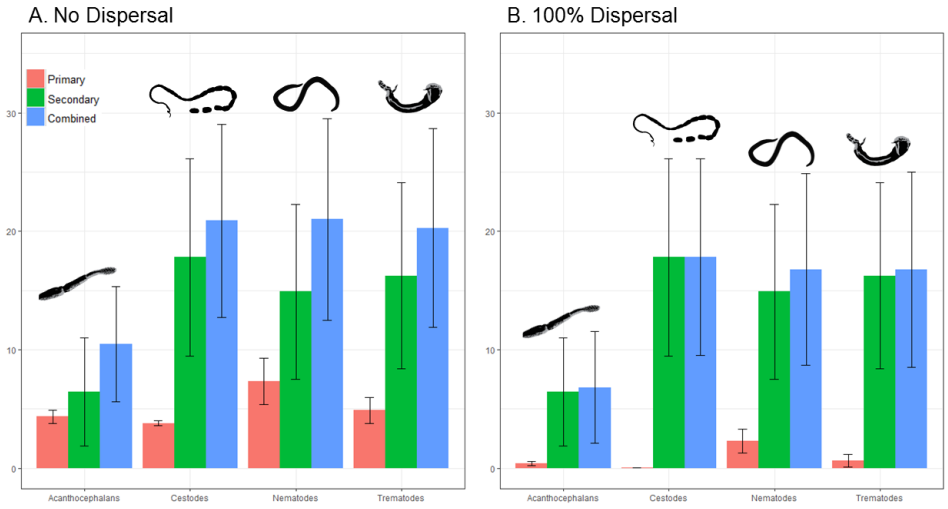


Figure 3.11: Primary, secondary and compounded extinction rates (%) for major helminth clades. Error bars represent lower and upper bounds to estimation based on the Thomas et al. method and errors in the Dobson method, and means between the two interval ends are shown in bars, for (left to right) acanthocephalans, cestodes, nematodes and trematodes. Cause of extinction is broken down by primary extinction (direct impacts of climate change, no dispersal), secondary extinction (coextinction with hosts, calculated in Appendix 1), and a combined risk (total). Scenarios are presented for (a) no dispersal and (b) full dispersal capacity for parasites. Most helminths face high risk when accounting for coextinction, though acanthocephalans consistently appear much less threatened.



Clade	*N*	**HL (Mean)**	**HL (5-95% CI)**	**% Comm. to Ext.**
0% Dispersal				
Acanthocephala	14	-16.6%	(-50.6%, -0.2%)	3.8% / 4.4% / 4.9%
Astigmata	18	-19.0%	(-43.6%, -4.0%)	4.4% / 5.1% / 5.3%
Cestoda	25	-13.6%	(-29.1%, -2.9%,)	4.0% / 3.6% / 3.7%
Ixodida	141	-31.9%	(-57.0%, -1.9%)	8.1% / 9.2% / 9.8%
Nematoda	147	-28.0%	(-74.4%, -2.6%)	5.4% / 7.9% / 9.3%
Phthiraptera	5	-55.8%	(-71.5%, -34.4%)	10.5% / 18.5% / 19.3%
Siphonaptera	67	-40.6%	(-69.5%, -11.0%)	10.0% / 12.2% / 12.9%
Trematoda	40	-17.8%	(-47.4%, -0.4%)	3.8% / 4.8% / 6.0%
100% Dispersal				
Acanthocephala	14	+48.8%	(-10.4%, +129.0%)	0.21% / 0.54% / 0.60%
Astigmata	18	+13.8%	(-41.2%, +64.4%)	1.3% / 2.0% / 2.3%
Cestoda	25	+57.1%	(+3.7%, +131.1%)	0.07% / 0.07% / 0.07%
Ixodida	141	-8.6%	(-54.1%, +67.7%)	4.9% / 5.7% / 6.4%
Nematoda	147	+18.7%	(-53.6%, +87.6%)	1.3% / 2.5% / 3.3%
Phthiraptera	5	+110.5%	(-57.7%, 514.8%)	4.6% / 6.2% / 7.4%
Siphon aptera	67	-5.0%	(-50.0%, +43.8%)	1.9% / 4.1% / 4.6%
Trematoda	40	+82.2%	(-30.4%, +242.4%)	0.11% / 1.0% / 1.2%

Table 3.1: Habitat loss and projected extinction risk by dispersal scenario and clade. Values are averaged across all General Circulation Models (GCMs) and Representative Concentration Pathway Scenarios (RCPs; 45), and the percent of species committed to extinction is calculated using the three Thomas *et al.* [8] SAR methods. Percentiles are calculated from species-level averages of GCMs and RCPs (i.e. all variance is interspecific).

Clade	N	**HL (Mean)**	**HL (5-95% CI)**	**% Comm. to Ext.**
0% Dispersal				
Acanthocephala	5	-19.8%	(-52.3%,0.0%)	5.2% / 5.4% / 6.2%
Astigmata (*)	2	-37.0%	(-44.1%,-0.3%)	9.2% / 10.9% / 11.0%
Cestoda	8	-20.4%	(-29.6%,-10.3%)	6.1% / 5.6% / 5.7%
Ixodida	98	-34.5%	(-55.3%, -2.6%)	8.5% / 10.0% / 10.7%
Nematoda	37	-24.2%	(-55.5%, -5.1%)	5.5% / 6.7% / 7.3%
Phthiraptera (*)	2	-57.9%	(-62.1%, -53.7%)	19.9% / 19.4% / 19.5%
Siphonaptera	30	-37.1%	(-51.7%, -16.7%)	9.4% / 10.9% / 11.3%
Trematoda	14	-23.8%	(-63.4%, -2.2%)	6.0% / 6.8% / 8.8%
100% Dispersal				
Acanthocephala	5	+38.2%	(-17.0%, +78.2%)	0.7% / 1.5% / 1.7%
Astigmata	2	-32.3%	(-43.1%, -21.5%)	6.8% / 9.3% / 9.6%
Cestoda	8	+30.6%	(-2.0%, +70.8%)	0.2% / 0.2% / 0.2%
Ixodida	98	-12.6%	(-53.3%, +47.6%)	4.9% / 6.0% / 6.6%
Nematoda	37	+13.6%	(-37.1%, +47.0%)	0.7% / 1.5% / 1.8%
Phthiraptera	2	-0.53%	(-60.1%, -45.9%)	17.9% / 17.2% / 17.4%
Siphonaptera	30	-0.03%	(-41.2%, +43.1%)	1.6% / 3.5% / 3.8%
Trematoda	14	+43.3%	(-24.9%, +138.4%)	0.2% / 1.2% / 1.7%

Table 3.2: Habitat loss and projected extinction risk by dispersal scenario and clade; analysis recreated from Table 3.1, but for 50+ point species.

Chapter 4

An ecological assessment of the pandemic threat of Zika virus.

Colin J. Carlson Eric Dougherty Wayne Getz

4.1 Abstract

The current outbreak of Zika virus poses a severe threat to human health. While the range of the virus has been cataloged growing slowly over the last 50 years, the recent explosive expansion in the Americas indicates that the full potential distribution of Zika remains uncertain. Moreover, many studies rely on its similarity to dengue fever, a phylogenetically closely related disease of unknown ecological comparability. Here we compile a comprehensive spatially-explicit occurrence dataset from Zika viral surveillance and serological surveys based in its native range, and construct ecological niche models to test basic hypotheses about its spread and potential establishment. The hypothesis that the outbreak of cases in Mexico and North America are anomalous and outside the native ecological niche of the disease, and may be linked to either genetic shifts between strains, or El Nino or similar climatic events, remains plausible at this time. Comparison of the Zika niche against the known distribution of dengue fever suggests that Zika is more constrained by the seasonality of precipitation and diurnal temperature fluctuations, likely confining autochthonous non-sexual transmission to the tropics without significant evolutionary change. Projecting the range of the diseases in conjunction with three major vector species (*Aedes africanus*, *Ae. aegypti*, and *Ae. albopictus*) that transmit the pathogens, under climate change, suggests that Zika has potential for northward expansion; but, based on current knowledge, our models indicate Zika is unlikely to fill the full range its vectors occupy, and public fear of a vector-borne Zika epidemic in the mainland United States is potentially informed by biased or limited scientific knowledge. With recent sexual transmission of the virus globally, we caution that our results only apply to the vector-borne transmission route of the pathogen, and while the threat of a mosquito-carried Zika pandemic may be overstated in the media,

other transmission modes of the virus may emerge and facilitate naturalization worldwide.

4.2 Introduction

Following a twenty-fold upsurge in microcephalic newborns in Brazil linked to Zika virus (ZIKV), the World Health Organization has declared an international health emergency. [239] Despite being profiled for the first time in 1947. [240] Zika remained poorly characterized at a global scale until the last six months. Thus, the present pandemic expansion in the Americas poses a threat of currently unknown magnitude. Closely related to dengue fever, Zika conventionally presents as a mild infection, with 80% of cases estimated to be asymptomatic. [241] The cryptic nature of infection has resulted in sporadic documentation of the disease and rarely includes spatially explicit information beyond the regional scale. [239, 242, 243, 244] This greatly limits the confidence with which statistical inferences can be made about the expansion of the virus. With an estimated 440,000–1,300,000 cases in Brazil in 2015, [241] and continuing emergence of new cases in Central America and, most recently, the United States, assessing the full pandemic potential of the virus is an urgent task with major ramifications for global health policy.

Current evidence portrays the global spread of ZIKV as a basic diffusion process facilitated by human and mosquito movement, a hypothesis supported by the frequency of infected traveler case studies in the Zika literature. [245, 246, 247, 248] Tracing phylogenetic and epidemiological data has revealed the expansion of ZIKV has occurred in a stepwise process through the South Pacific, moving the disease from Southeast Asia into French Polynesia and the Philippines, and subsequently to Easter Island. [239, 242, 243, 244] Based on phylogenetic reconstruction, ZIKV is assumed to have dispersed into South America as recently as three years ago from the last of those locations, [249] and the virus is presumed to be at a biogeographic disequilibrium in the Americas. With cases in the ongoing outbreak in Colombia, El Salvador, Guatemala, Paraguay, and Venezuela, and by November of last year, as far north as Mexico and Puerto Rico, the full potential distribution of the disease remains unknown. Significant fear exists regarding the potential naturalization of Zika across the United States, given a handful of sexually-transmitted examples, but as of now the disease has not been carried into the United States by vectors. Moreover, several alternative explanations for the diseases expansion remain overlooked; most notably, the role of climate change in Zikas expansion has not yet been thoroughly investigated. [250]

We present three competing hypotheses that describe the path of expansion that Zika could take, based on evaluations of the ecological niche of the virus within and outside of its vectors. First, if Zika has no additional climatic constraints relative to those of its vectors, future range expansions should match mosquito ranges. Second, if Zika has a transmission niche that is constrained by climatic factors within the ranges of its mosquito vectors, its range may be much more limitedwith, as we show below, possible confinement to the tropicsand cases in North America could be driven by human dispersal or extreme episodic weather events. Finally, it is possible that the expansion of Zika into North America may be a steady

range expansion beyond the known niche in its native range, facilitated by climatic shifts or by genetic shifts in the virus or vectors. To test these hypotheses, we present a spatially explicit database of Zika occurrences from the literature and an ensemble of ecological niche models [251] using that data to map the potential distribution of the virus.

4.3 Methods

Occurrence Data

Occurrence data for Zika virus was compiled from the literature from studies dating as far back as the original discovery of the virus in Zika Forest, Uganda in 1947. While the asymptomatic nature of the virus limits the total availability of data, lack of evidence for spatial patterns in symptoms in the native range suggest this is an unlikely cause of spatial bias (and instead, merely limits total dataset size). Special attention was paid to correctly attributing cases of travelers to the true source of infection. Locality data was extracted from a combination of clinical cases and seropositivity surveys in humans and mosquitoes, and georeferenced using a combination of Google Maps for hospitals and the Tulane University GEOLocate web platform for the remainder, [164] which allows for the attribution of an uncertainty radius to points only identified to a regional level. To our knowledge, this spatially explicit database is the most inclusive dataset currently in the literature. Four points were georeferenced in the New World but excluded from niche models because a limited sample as small as four points was likely to significantly bias predictions (compared to the necessary number of pseudoabsences in the same region). Thus, sixty points from the Old World were used in the final models presented in our paper after eliminating data from the current outbreak in the Americas. All points included in our dataset had an outer-bound of at most 65 km of uncertainty, with most substantially less.

Constraining datasets based on an uncertainty threshold will become more statistically feasible in future studies once more survey data become available. In the present study, we deemed that the additional information gained from each point outweighed the potential impact of the uncertainty on model performance (The dataset is available at `https://figshare.com/articles/An_Ecological_Assessment_of_the_Pandemic_Threat_of_Zika_Virus/3790698`). We note that for similar reasons, we did not subsample our dataset for spatial thinning in our main models, as software packages like spThin allow, [252] due to information-accuracy tradeoffs; and the strong final performance of models (and the correspondence of our predictions for dengue and *Aedes* species to published gold standard niche models) speaks to the appropriateness of the underlying data and variables. Sensitivity analyses in the literature unequivocally suggest that accuracy of the modeling methods we employ plateaus at or near 50 points, justifying the use of a dataset of this size. [166, 167, 168]

Occurrence data for the other species included in our study were compiled from the literature. For *Aedes africanus*, we used a dataset of 99 points downloaded from the Global

Biodiversity Informatics Facility (www.gbif.org). GBIFs coverage of *Aedes aegypti* and *Aedes albopictus* was deemed to be lacking, so occurrences for those species were taken from the previously published work of Kraemer *et al.* [253, 254] Finally, Messina *et al.*s database was used for dengue, [255] as it has been previously published used with great success to generate a global distribution model. [256] Both of these datasets were reduced down to point-only data (i.e., polygons of occurrence were excluded), leaving 5,216 points for dengue and 13,992 and 17,280 points for *Ae. aegypti* and *Ae. albopictus* respectively.

A number of other Zika vectors are known from previous reports, including at least a dozen *Aedes* species, as well as *Anopheles coustani*, *Culex perfuscus*, and *Mansonia uniformis*. [257, 258] While we do not include these vectors in this study in order to keep focus on the most likely globally-cosmopolitan *Aedes* vectors, we note these species could be important in regional patterns of establishment. These species lack the globally comprehensive datasets that dominant arbovirus-vectoring *Aedes* species have, and require future attention by similarly-dedicated researchers.

Ecological Niche Modeling

Due to the potentially transient nature of the New World distribution of Zika virus, our model uses presence and 1000 randomly selected pseudo-absence points from the Eurasian, African, and Australian regions where the virus is established. We used the WorldClim data set BIOCLIM at 2.5 arcminute resolution, an aggregated dataset across values from 1950 to 2000, to provide all but one of our climate variables. [259] The BIOCLIM features 19 variables (BIO1-BIO19) that summarize trends and extremes in temperature and precipitation at a global scale. Given the relevance of the normalized difference vegetation index (NDVI) in previous studies of dengue and as a predictor of vector mosquito distributions, [260] we downloaded monthly average NDVI layers for each month in 2014 from the NASA Earth Observations TERRA/MODIS data portal, [261] at a resolution of 0.25 degrees to maintain compatibility with the BIOCLIM layers (0.25 degrees is equivalent to 15 arcminutes). The twelve monthly layers were averaged to provide a single mean NDVI layer. Due to the absence of NDVI data at the necessary resolution associated with many of the historical records (especially prior to 1992), the use of a recent mean NDVI layer was deemed the most pragmatic method of including vegetation in our models. We also make the simplifying assumption that areas of prior presence correspond to areas of current presence, an assumption that allows the use of current NDVI and is relatively standard for the niche modeling literature.

Species distribution models were executed using the BIOMOD2 package in R 3.1.1, which produces ensemble species distribution models using ten different methods: general linear models (GLM), general boosted models or boosted regression trees (GBM), general additive models (GAM), classification tree analysis (CTA), artificial neural networks (ANN), surface range envelope (SRE), flexible discriminant analysis (FDA), multiple adaptive regression splines (MARS), random forests (RF), and maximum entropy (MAXENT). [262] The BIOMOD algorithm runs a series of distribution models using training data, each of

which is subsequently weighted and stacked across methods based on relative predictive performance with test data. As Thuiller *et al.* note, if a single modeling method is consistently most accurate, use of that method should be favored over ensemble approaches, [262] but in our study model performance varied, making ensemble approaches informed by degree-of-belief in a given model the most powerful option available. With recent publication of two Zika niche modeling papers using MAXENT and boosted regression trees, respectively, [12, 11] differences between these two modeling methods may be responsible for differences in predictions an issue that makes ensemble models particularly robust to idiosyncrasies of any individual methods.

Models were run individually for Zika (ZIKV), dengue (DENV), Ae. aegypti, Ae. albopictus, and Ae. africanus. For Zika, models trained on Old World environmental data (from Europe, Africa, Asia and Australia) were used to establish the potential distribution of the virus in the Americas under climatic conditions captured by WorldClim data, which are an aggregate of data between 1950 and 2000 (appropriately matching the date range of historical Zika occurrence data), and represent an expected range of variability that does not incorporate anomalous events like 2015 El Nio Southern Oscillation. Extrapolation between continents is a procedure with the potential for error: if novel environments exist in the New World with incomparable covariance structure between climate variables, predictive accuracy is likely to decline. While using only Old World data could potentially bias our models towards a subset of the niche, this can be readily tested for, by comparing models that include or exclude South American occurrence data.

To address colinearity in the environmental variable set, we produced a correlation matrix for our 20 variables, and identified each pair with a correlation coefficient ¿ 0.8. For each species, we ran a single ensemble model with all ten methods and averaged the variable importance for our 20 predictors across the methods (Table 4.1-4.5). In each pair we identified the variable with the greater contribution, and we produced species-specific reduced variable sets used in the final published models by eliminating any covariates that universally performed more poorly than their pair-mate. Based on this criterion, we excluded the following variables for each species to reduce colinearity:

- ZIKV: BIO8, BIO9, BIO14, BIO18
- DENV: BIO3, BIO5, BIO12, BIO17
- *Ae. aegypti*: BIO6, BIO8, BIO12, BIO17
- *Ae. africanus*: BIO5, BIO6, BIO12, BIO17
- *Ae. albopictus*: BIO8, BIO9, BIO16, BIO17

The AUC of every model run with reduced variable sets is presented in Table 4.6. We found no significant correlation between NDVI and any individual BIOCLIM variable, so NDVI was included in every model of current distributions. We ran five iterations of each

reduced variable set model and eliminated any prediction methods from the ensemble with an AUC of lower than 0.95, so that the final model had only included the best predicting models. This greatly limited the models available for ZIKV and DENV, so a cutoff of 0.9 was applied in those cases, to keep the ensemble approach constant across datasets. The final models were run with the following methods with ten iterations using an 80/20 training-test split in the final presentation:

- ZIKV: GLM, GBM, GAM, CTA, FDA, MARS, RF
- DENV: GLM, GBM, GAM, FDA, MARS, RF, MAXENT
- *Ae. aegypti*: GLM, GBM, GAM, CTA, ANN, FDA, MARS, RF
- *Ae. africanus*: GLM, GBM, GAM, CTA, ANN, FDA, MARS, RF
- *Ae. albopictus*: GLM, GBM, GAM, CTA, FDA, MARS, MAXENT, RF

The importance of variables of the reduced model set for each are presented in Table 4.7-4.12, and the final ensemble models are projected from the BIOMOD output in Figure 4.6-4.11.

Model Validation

To assess the transferability of our Zika model across environmental space, we conducted a geographic cross validation (GCV) between African and Asian datasets (an analysis we did not repeat for *Aedes* species or dengue, given the far greater sample size and geographic coverage of those species, and the publication of more intensive niche modeling efforts by experts for those systems). While under normal circumstances, a model would be trained on New World data and projected onto the Old World to cross-validate results, the lack of data prior to the current outbreak makes such a direct comparison infeasible. However, given the evidence for separate Asian and African strains, a cross-validation between the two was supported, and models trained on those two continents were projected globally to test the performance of the model across geographic regions, and evaluate how sensitive our projections in the Americas are to the environmental covariates sampled. The clustering of points in western India narrows the environmental range sampled by presences, potentially limiting the apparent transferability of the Asian sub-model. In contrast, the African sub-model performs well in new regions, and corresponds well to the global model.

Climate Change Projections

The potential contribution of climate change to Zikas current expansion, and the outer bounds of transmission under future expansion, are largely unaddressed. While these have not been the subject of any concerted speculation, Shapshak *et al.* [263] point out that the majority of arboviruses are potentially implicated in the climate change-driven expansion of

global disease burden, with a shared set of drivers that quite probably extends to Zika as well. Consequently this analysis serves two purposes; to address the potential expansion and thereby assist public health planning, and to test whether even a liberal post-climate-change interpretation of range margins matches the predictions of Messina *et al.* [12] and Samy *et al.* [11] that we consider limited in specificity and potentially over-predictive. To project the distribution of the species under a worst-case scenario for climate change, we reran each model with the previously chosen method and variable sets but excluded NDVI, as future values could not be simulated effectively. BioClim forecasts were taken from WorldClim using the Hadley Centre Global Environmental Model v. 2 Earth System climate forecast (HadGEM2-ES) predictions for representative climate pathway 8.5 (RCP85), which, within that model, represents a worst-case scenario for carbon emissions and climate warming. [264] All five species models were retrained on current climate data and projected onto forecasts for the year 2050. While we could have also included milder climate change forecasts and scenarios in our analysis, public concern over the future spread of Zika make the worst case scenario the most relevant question of interest for public health research (and intermediate scenarios would fall between current ranges and the worst case scenario we project).

Niche Comparison

To compare the niche of dengue and Zika and thereby address whether dengue models can be appropriately used to forecast the Zika pandemic, we used the R package ecospat, which uses principal component analysis to define the position of species ecological niche relative to background environmental variation. [265, 266] The ecospat analysis was run using the full 64 point database and the full extent of global environmental data, because, while the niche of Zika in the Americas is uncertain, dengue is well established, and the analysis was most appropriately done with global coverage. Niche similarity tests were run with 500 iterations and using the entire set of 20 environmental variables (Bioclim + NDVI).

Model Comparison with Global Data Coverage

Our study is centered on the assumption that incorrect predictions at the country level can have drastic consequences for the misinterpretation of science. As a final precautionary analysis, we supplemented the data published in the Messina *et al.* study [11] to our own for a final re-analysis. Broennimann & Guisan [267] recommend the pooling of data from native and invasive ranges for ecological niche modeling during the course of a biological invasion, an approach we adopt in this final analysis. The Messina data is heavily clustered in Brazil, with a high degree of aggregation, and especially compared against our less-aggregated, smaller dataset this made the combination of datasets potentially inaccurate. To address this problem, the 390 pooled points were reduced down to 242 points using the package spThin, [252] with a 40km buffer between points (the width of an average grid cell for our environmental data). Models were rerun using the same variable and model set as for the primary Zika model and the results of the analysis are included in the supplementary

information as Figure 4.11 and, with a threshold applied based on the true skill statistic, Figure 4.12. The final model performs poorer than our main ensemble (weighted model: AUC = 0.970), and while it more appropriately predicts presences in southern Brazil, it does a far poorer job in the rest of the world, once again most likely due to the relative balance of points even after thinning the dataset.

4.4 Results

Our final Zika model combines seven methods with a variable set chosen from bioclimatic variables and a vegetation index to minimize predictor covariance. The ensemble model performs very well (AUC = 0.993; Figure 4.1), to a degree that resembles overfitting but is in fact driven by the strength of the ensemble modeling approach (which preferentially weights the best models across iterations, minimizing the error associated with any given high-performing iteration). The model strongly matches most occurrences including the hotspots of Brazilian microcephaly. It also predicts additional regions where Zika is so far unrecorded, but where further inquiry may be desired (in particular, Southern Sudan and the northern coast of Australia). Our model indicates that certain occurrences, like the 1954 report from Egypt and almost all North American cases, are likely outside the stable transmission niche (i.e., persistent over time) of the virus (sensu [268]). Moreover, we note that visual presentation of cases or, of ecological niche models at the country level may make the range of the virus appear far larger than our models suggest (see Figure 4.1).

Given the public health crisis posed by Zika, and the potential costs associated with underpredicting the extent of the current outbreak, we pay special attention to evaluating the sensitivity of our models to variations in our preliminary dataset. Historical geographical data on cases in the Americas are lacking, given the recent introduction of the virus, and the routes and drivers of transmission involved in that outbreak are uncertain, preventing meaningful cross-validation of models of the current outbreak with our Old World model. However, it is worth noting that recent phylogenetic work suggests a deep phylogenetic division between African and Asian strains, the latter of which as a monophyletic group include the entire radiation through French Polynesia into current outbreak areas; [249, 269] to address the potential evidence that African and Asian strains of the virus may be ecologically distinct, we present models trained on each continent and projected globally as a basic sensitivity analysis (Figure 4.2).

The two models cross-validate weakly compared to the performance of the global model; driven by both the 50% reduction in sample size and the higher degree of aggregation of Asian occurrences, the two projected distributions are dramatically different. Despite the over-prediction of the Asian model in Africa and the possible overfitting of the African model, we emphasize that neither extreme scenario predicts any substantially greater range in North America than our main ensemble model. Moreover, as projected in North America, our Asian model underpredicts but does predict two major hotspots of occurrence in Brazil, the Ceara/Rio Grade do Norte region and Roraima, both of which spatially corre-

spond to hotspots of Zika according to the recent Faria *et al.* publication in Science, [249] adding further support to the model. Finally, despite low transferability between continents, both sub-models are well matched by our aggregated model in their native range, further supporting the accuracy and predictive power of our global projection.

Recently published work by Bogoch *et al.* [270] uses an ecological niche model for dengue as a proxy for the potential full distribution of ZIKV in the Americas, presenting findings in terms of potential seasonal vs. full-year transmission zones. While that approach has been effectively validated for dengue transmission in mosquitoes, using a model of one disease to represent the potential distribution of another emerging pathogen is only a placeholder, and is particularly concerning given the lack of evidence in our models that ZIKV and dengue have a similar niche breadth. [271] Comparing our niche models for dengue and ZIKV reveals that the two niches are significantly different (Schoener's D = 0.176; p ¡ 0.01; Figure 4.3). While the two occupy a similar region of global climate space, Zika is more strictly tropical than dengue, occupying regions with higher diurnal temperature fluctuations and seasonality of precipitation (Figure 4.3a).

Projecting niche models to the year 2050 suggests that expansion of Zikas niche outside the tropics is an unlikely scenario, independent of vector availability (Figure 4.4). However, significant westward expansion in South America and eastward expansion in Africa implies that Zika may continue to emerge in the tropics. Moreover, our future projections for dengue (which strongly agree with previously published ones [272]) show an expansion out of the tropics that is not shared with Zika (Figure 4.4). These results call into question the applicability of dengue niche models used to project a significant future range for Zika in North America. [270]

Finally, we add a last layer of validation in the form of an analysis aggregating our and Messina *et al.*s data, and include the results of an updated ensemble model in Figure 4.5 (as well as 4.6 and 4.7). Even with spatial thinning, that updated model is still heavily biased in favor of the South American occurrence data, which it predicts excellently, compared to a weaker fit in Africa and Asia. That accompanying loss of specificity is partly responsible for a lower AUC than our main model (AUC = 0.970) and the low TSS-based threshold (271, from 0 to 1000) that produces the substantially-greater predicted range shown in Figure 4.12. The model does predict the current outbreak more effectively than ours, in particular better encompassing the southern half of Brazil where a surprising number of cases are clustered. But those southward expansions are accompanied by far less expansion above the equator in the Americas, and once again with the exception of the southernmost tip of Florida, there is no substantial predicted range in the United States, even along the Gulf Coast. If model discrepancies are attributed to evolutionary change and not to differences in model methods and specificity, those evolutionary changes seem to have done little to expand the North American niche of the virus.

4.5 Discussion

Ecological niche modeling has become one of the most generalized and useful parts of the streamlined response process for emerging infections. Recently published ecological niche models for Zika using MAXENT [11] and boosted regression trees [255] have resulted in somewhat conflicting results. Samy *et al.*, using data exclusively from the range of the current outbreak, project autochthonous transmission in the southeastern United States, and potentially throughout the U.S. following regional outbreaks introduced by travelers. Their analysis incorporates socioeconomic factors into prediction, a valuable extra dimension we did not incorporate into our analysis; but the prediction of regions throughout the United States and most of the European continent as suitable based on only these criteria (i.e. despite lacking available vectors) seems uninformative except for the prediction of sexual outbreaks. Samy *et al.*, however, conclude: In Western Europe, ZIKV transmission risk is enhanced by travel times and connectivity to known transmission areas; as such, isolated autochthonous cases may occur at least seasonally when competent vector species are present. [11] Messina *et al.* have a similar finding, based on a primarily ecological approach applied to 323 occurrences mostly from the New World; they map out most countries in the world as highly suitable, including the United States, with the conclusion that 2.17 billion people live in countries within Zikas potential expanse. [12] These studies, being contemporaneous, do not refer to each other, and their conflicting results could render Zika forecasts unclear to the media and policymakers.

Interpreting conflicts between these models and those published here requires acknowledging three fundamental problems. First, differences in virulence between American and Asian strains of the virus may have changed the range limits. The niche of the vector-borne disease is manifest in its transmission and prevalence in mosquitoes (as well as humans and reservoirs), and increases in virulence could change the threshold of habitat suitability manifest in range limits. Without comparative work using updated data in Samy *et al.* and Messina *et al.*s papers, equal support exists for our differences being attributable to methodological discrepancies or to a difference between Asian and American strains. But in the preliminary analysis we present in the supplementary information, incorporating data from the New World does not substantially expand projections in the United States (though a greater region of Brazil is predicted); and we believe a combination of evolutionary shifts and methodological differences is likely the most parsimonious explanation for differing results.

Second, we acknowledge the untested possibility that Zika has been expanding in its range since discovery in the 1940s (though, the virus was soon recorded in Borneo and Vietnam in the 1950s [257]), which would also decrease both the accuracy of our models in that region, and their power in the New World compared to the models published in the other two studies. Testing that possibility using our data broken down by time periods would be strongly statistically biased by the non-random element of viral discovery in different tropical countries, a factor for which it would be nearly impossible to control. Phylogenetic evidence has placed the introduction in the Americas within the last decade [249], but the age of divergence between Zika and closely related viruses like Japanese and St. Louis

Encephalitis Viruses is less certain. Improving phylogenetic evidence based on updated Old World genomes in the coming years is a far more appropriate methodology for testing different biogeographic theories within that region.

Third and finally, we acknowledge the possibility that dispersal limitations have changed between the Old and New World, in such a way that the present expansion of Zika is not the emergence of novel niche space but the manifestation of hidden plasticity. This possibility is troubling from a public health perspective: if Zikas niche is simply more expansive than current data/models capture, its geographic expansion could progress much further than we predict. This problem is fundamental to all predictive models applied to biological invasions, but Broenniman & Guisan [267] suggest that combining data from the native and invasive range maximizes the utility of ENMs in these scenarios. In our combined model we find evidence for subtle differences, especially in South America, but our findings remain sound with respect to the boundaries of transmission in North America. In any niche modeling study, there is always the possibility for error by omission; but we find no evidence that this has occurred in our study.

The dynamics of arboviruses at the range margins of their vectors are complex. In the case of dengue, the distribution of the virus in the United States (and elsewhere in temperate regions) remains more constrained than the range of its vectors. Our paper tests and rejects the hypothesis that predictions of Zika will occupy the entire niche of Aedes populations in North America, disagreeing with the two recently published niche model studies. Our models imply a similar constraint on Zika transmission to that of dengue if not a more pronounced one, and owing to the complexities surrounding transmission dynamics at the edges of suitable ranges, [273] the potential existence of Zika in even the southernmost parts of Florida [274] may not sustain autochthonous Zika transmission indefinitely. Making more specific predictions within Florida can be done through ecological niche models, but is likely more appropriately achieved through conventional epidemiological models that explicitly model vector abundance, biting rates and phenology.

Our models find an ecological nonequivalence of Zika and dengue, and suggest that the niche of the virus in both Africa and Asia is far narrower than what other models project based on current outbreak data or based on knowledge of dengues spread. We reject our first hypothesis, but based on the occurrence of Zika cases outside our predicted suitable range for the virus, we cannot eliminate our second hypothesis that the 2016 Zika outbreak may be in ephemeral, rather than stable, parts of the Zika transmission niche due to episodic climatic conditions. Specifically, El Nio Southern Oscillation (ENSO) events drive outbreaks of dengue in the Americas and in Southeast Asia, [275] and Paz *et al.* [250] have conjectured that the 2015 ENSO event could have contributed to the severity of the ZIKV outbreak in North and Central America (in response to Bogoch *et al.* [270]). While wind-dispersed mosquitoes carrying infections can be responsible for the introduction of diseases to new regions, [276] reported cases in the United States have all been contracted sexually or while traveling abroad to regions with endemic outbreaks, further supporting the tropical constraint hypothesis. However, in the second hypothesis scenario, the rapid expansion during the current outbreak beyond the boundaries of the stable transmission

niche is unlikely to be followed by naturalization of the pathogen in the United States in the future, except perhaps in the southernmost tip of Florida. While ecological niche models relate occurrence to climate, drivers of disease may operate at the temporal scale of weather, and we suggest further analyses of a different methodology are necessary to confirm or reject the potential contribution of El Nino or anomalous storms to Zikas expansion.

In the case of our third hypothesis, if alternative modeling efforts based on data from the Americas are evidence that the niche of the American strain of the virus has broadened, it is possible that mutations allowing increased virulence or changing transmission dynamics have occurred (and that weather events have not driven the severity of the current outbreak). From the results of our supplementary analysis using aggregated global data, we continue to treat the third hypothesis as a hypothesis for which there may be weak evidence. But we suggest it cannot be rejected or accepted confidently unless alternative hypotheses are eliminated and more evidence is collected in particular, empirical data demonstrating or failing to find differences in transmission dynamics or virulence between the native Asian virus and its invasive descendant (rather than global comparisons and cross-validations of different ecological niche models).

Our models nevertheless suggest it could be premature to expect Zika naturalization as a widespread eventuality in North America, as other models have forecasted. Without more definitive information on the basic biology of Zika, however, the confidence with which niche models can forecast pandemics is limited. In particular, we also draw attention to recent evidence suggesting Zika persistence may depend on wildlife reservoirs in addition to human hosts and mosquitoes. Primates have been suggested as the primary candidate clade because the Zika flavivirus was first isolated in a rhesus macaque in the Zika Forest in Uganda. But as rhesus macaques do not occur on the African continent, and were captive there for inoculation experiments, the primate reservoir hypothesis remains unsupported. A 2015 case of an Australian presumed to have contracted Zika from a monkey bite while traveling in Indonesia, however, indicates that primates may transmit the virus directly. [247] Additionally, antibodies against Zika have been observed in several rodent and livestock species in Pakistan, [277] as well as several large mammal species, including orangutans, zebras, and elephants. [278] The potential for any North American wildlife species to play host to Zika is, at the present time, entirely unknown, and the emergence of novel amplification hosts (which may allow the virus to proliferate above the host density threshold in vectors in regions otherwise unsuitable for sustained transmission) could potentially expand the suitable range margins of Zika infection on a global scale.

From the results of our model we find strong evidence for the hypothesis that the global threat of a specifically vector-borne Zika pandemic, though devastating, may be most acute in the tropics; and we find that the evidence of future North American transmission in the literature is not unequivocal. However, we concur with the scientific majority that sexual transmission of Zika infections may still facilitate a significant outbreak in the United States and other previously unsuitable regions, particularly under evolutionary processes that select for the most directly transmissible strains of pathogens. [279] A case of sexual transmission in Texas has been suspected in the 2016 outbreak, and two previous reports of likely sexual

transmission of ZIKV occurred in 2011 and 2015. [243, 280] Even if the Zika cases in the United States represent a rare spillover outside of the mosquito-borne viral niche, sexual transmission could create a new, unbounded niche in which the virus could spread. We draw attention to the potential parallels with simian and human immunodeficiency virus (SIV/HIV), for which a sexually transmitted pandemic has overshadowed the zoonotic origin of the disease. [281] With Zikas asymptomatic presentation and the overall confusion surrounding its basic biology and transmission modes, we caution that its potential for severe sexually-transmitted outbreaks cannot be overlooked in the coming months.

To address the broader community of modelers and ecologists involved in the Zika intervention, we conclude with a final cautionary note. The consequences of under-predicting an outbreaks potential distribution are obvious and our results are phrased cautiously as a result. But there are also economic and social consequences to over-predicting the potential distribution, especially in the United States. The response to Zika is necessarily political and consequently involves the division of resources between domestic preparedness and international relief; while new tools are being developed to help allocate funds efficiently based on epidemiological principles (we particularly highlight the work of Alfaro-Murillo *et al.* [282]), global overestimation of the viruss trajectory could vastly reduce the power of those methods.

Models like those of Messina *et al.* and Samy *et al.* that predict substantial Zika expansion in the United States, and in the case of the former suggest Zika could threaten up to 2.17 billion people, contribute (independent of accuracy) to fear of an American pandemic. This prediction necessarily diverts funding away from relief efforts in Brazil and other affected countries in Latin America, increasing the probability of traveler infections feeding sexual outbreaks in the U.S.; and further reduces the credibility and impact of the American foreign response to Zika by mobilizing potentially-unnecessary domestic responses. At the time of writing, the Zika Vector Control Act passed by the U.S. House of Representatives weakens permit requirements for spraying pesticides near bodies of water without reallocating any funding for Zika interventions; and preventative efforts in New York City alone will cost $21 million to trap mosquitoes and hire epidemiological experts, with other cities outside our predicted range investing in preparation and vector control to similar degrees. Voices of scientific authority contributing to fear in the United States can substantially impact the political response to Zika, and it serves future modeling efforts to be as accurate, cautious, and objective as possible in the information and statistics that underpin media and policy conversations. But even more importantly, scientific teams with different approaches and data must work collaboratively to interpret the discrepancies between their results and to build an unbiased scientific consensus that is accessible to the public.

4.6 Acknowledgements

C.J.C. thanks Fausto Bustos for feedback on initial ideas presented in the paper, and Kevin Burgio for extensive methodological support, training, and mentorship. We also thank four

anonymous reviewers for their aid in strengthening the manuscript.

Table 4.1: Zika full variable set preliminary model variable importance

	GLM	GBM	GAM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT
bio1	0.573	0.009	0.727	0	0.089	0.385	0.001	0	0.022	0.294
bio2	0.896	0.006	0.491	0	0.072	0.361	0	0	0.013	0.546
bio3	0.841	0.015	0.46	0	0.14	0.286	0.056	0	0.025	0.215
bio4	1	0	0.552	0	0.527	0.372	0	0	0.014	0.06
bio5	0	0.003	0.659	0	0.421	0.231	0	0	0.014	0.487
bio6	0.611	0.205	0	0.935	0.218	0.463	0.468	0.708	0.064	0.177
bio7	0	0.031	0.444	0	0.62	0.296	0.151	0.343	0.04	0.263
bio8	0.219	0	0.439	0	0.211	0.412	0	0	0.007	0
bio9	0	0.009	0.625	0	0.012	0.302	0	0	0.014	0
bio10	1	0.002	0.698	0	0.036	0.336	0	0.266	0.012	0
bio11	0	0.019	1	0	0.155	0.378	0	0	0.059	0.076
bio12	0	0.019	0.191	0	0.453	0.106	0.188	0.348	0.029	0.126
bio13	0	0.092	0.204	0	0.759	0.211	1	0.738	0.066	0.534
bio14	0.589	0.003	0.377	0.211	0.303	0.013	0.514	0.065	0.003	0.302
bio15	0	0.047	0.201	0.558	0.001	0.068	0.073	0.125	0.017	0.533
bio16	0	0.009	0.18	0	0.537	0.155	0.618	0.786	0.028	0.031
bio17	0.588	0.003	0.501	0	0.345	0.018	0.572	0	0.008	0.464
bio18	0	0.002	0.044	0	0.331	0.145	0.054	0.184	0.006	0.178
bio19	0	0.008	0	0	0.173	0.059	0.114	0.286	0.007	0.307
NDVI	0	0.073	0.272	0	0.025	0.27	0.043	0.204	0.048	0.624

Table 4.2: Dengue full variable set preliminary model variable importance.

	GLM	GBM	GAM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT
bio1	0.295	0.001	0.366	0.052	0.04	0.242	0	0	0.015	0.007
bio2	0	0.035	0.136	0.202	0.016	0.1	0.078	0.004	0.044	0.048
bio3	0.062	0	0.054	0.065	0.002	0.216	0.018	0	0.016	0.05
bio4	0.21	0.013	0.59	0.024	0.262	0.241	0	0	0.03	0.013
bio5	0.146	0	0.293	0.011	0.024	0.196	0	0.029	0.008	0.008
bio6	0.172	0.013	0.286	0.41	0.08	0.271	0.348	0.5	0.035	0.011
bio7	0.427	0.031	0.724	0.161	0.134	0.232	0.057	0.298	0.045	0.038
bio8	0.203	0.01	0.152	0.041	0.017	0.262	0.004	0.315	0.015	0.023
bio9	0.242	0	0.149	0.005	0.084	0.257	0	0	0.007	0.024
bio10	0.477	0	0.521	0.002	0.006	0.209	0.036	0.078	0.007	0.003
bio11	0.94	0.068	0.639	0.005	0.002	0.274	0.797	0.352	0.043	0.212
bio12	0.033	0	0.095	0.024	0.059	0.124	0.141	0	0.016	0.009
bio13	0.324	0.007	0.083	0.017	0.052	0.207	0.714	0.472	0.013	0.003
bio14	0.111	0.004	0.31	0.047	0.008	0.017	0.048	0.348	0.025	0.078
bio15	0.003	0.001	0.075	0.016	0.018	0.079	0.017	0	0.007	0.022
bio16	0.118	0	0.113	0.015	0.208	0.16	0	0.105	0.01	0.025
bio17	0.033	0.004	0.102	0.102	0.13	0.074	0	0.172	0.022	0.006
bio18	0.013	0.002	0.007	0.005	0.031	0.117	0.004	0	0.017	0.015
bio19	0.012	0.018	0.035	0.048	0.052	0.079	0.056	0.029	0.024	0.073
NDVI	0.073	0.032	0.075	0.105	0.016	0.198	0.067	0.039	0.035	0.099

Table 4.3: *Aedes aegypti* full variable set preliminary model variable importance

	GLM	GBM	GAM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT
bio1	0.288	0.003	0.416	0.003	0.032	0.258	0.209	0	0.009	0.013
bio2	0.111	0.011	0.029	0.242	0.023	0.165	0.023	0.003	0.054	0.017
bio3	0.319	0.004	0.136	0.004	0.01	0.235	0.301	0.218	0.013	0.096
bio4	0.931	0.002	0.227	0.015	0.131	0.254	0.486	0	0.014	0.052
bio5	0.249	0	0.235	0	0.066	0.251	0.298	0	0.011	0.015
bio6	0	0.001	0.056	0.026	0.017	0.256	0	0	0.017	0.05
bio7	0.644	0.011	0.123	0.052	0.155	0.249	0.048	0.078	0.036	0.026
bio8	0.027	0.034	0.027	0.081	0.041	0.272	0.077	0.21	0.028	0.034
bio9	0.106	0	0.028	0.033	0.073	0.249	0.021	0	0.007	0.027
bio10	0.953	0	0.617	0.006	0.029	0.262	0.785	0	0.01	0.016
bio11	0.221	0	0.605	0.016	0.029	0.245	0	0	0.011	0.05
bio12	0.008	0.001	0.004	0.049	0.173	0.248	0	0.009	0.011	0.028
bio13	0.364	0.01	0.093	0.072	0.031	0.252	0.416	0.338	0.016	0.006
bio14	0.045	0.002	0.056	0.013	0.016	0.032	0.094	0.034	0.013	0.026
bio15	0.009	0	0.012	0	0.073	0.106	0	0	0.009	0.077
bio16	0.064	0	0.05	0.007	0.099	0.251	0.057	0.035	0.016	0.006
bio17	0.033	0	0.031	0.009	0.048	0.117	0.065	0	0.008	0.009
bio18	0.076	0.099	0.054	0.141	0.431	0.176	0	0.006	0.04	0.23
bio19	0.006	0.003	0.009	0.059	0.03	0.145	0.015	0.012	0.014	0.016
NDVI	0.044	0.019	0.029	0.029	0.059	0.245	0.042	0.042	0.016	0.021

Table 4.4: *Aedes africanus* full variable set preliminary model variable importance

	GLM	GBM	GAM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT
bio1	0.437	0.003	0.56	0.245	0.141	0.305	0.231	0.085	0.005	0
bio2	0	0.006	0.54	0.217	0.051	0.055	0	0	0.005	0.251
bio3	0	0.009	0.628	0	0.049	0.465	0	0	0.029	0.489
bio4	0.032	0.372	0.751	0.845	0.726	0.486	0	0.56	0.04	0.42
bio5	0	0	0.564	0	0.089	0.235	0	0	0.002	0.065
bio6	0	0.003	0.477	0	0.137	0.317	0.09	0.368	0.015	0.132
bio7	0	0.012	0.73	0.108	0.165	0.367	0	0.406	0.012	0.253
bio8	0	0.008	0.603	0	0.186	0.336	0.126	0.402	0.01	0
bio9	0.625	0.003	0.617	0	0.108	0.311	0	0.388	0.006	0
bio10	0.447	0.001	0.709	0	0.133	0.252	0	0.419	0.003	0.183
bio11	0	0.025	0.605	0.094	0.365	0.457	0.953	0.629	0.019	0.248
bio12	0.11	0.007	0.484	0.03	0.185	0.386	0.02	0.266	0.007	0.043
bio13	0.231	0.002	0.381	0	0.344	0.399	1	0.43	0.007	0.118
bio14	0.177	0.001	0.3	0	0.112	0.035	0.033	0	0.004	0.35
bio15	0.243	0.001	0.349	0	0.035	0.148	0	0	0.003	0.128
bio16	0	0.004	0.478	0	0.513	0.323	0.532	0.431	0.008	0.267
bio17	0	0.002	0.074	0	0.103	0.04	0	0	0.005	0.117
bio18	0.28	0.011	0.307	0.186	0.186	0.23	0.085	0.099	0.007	0.397
bio19	0.293	0.124	0.446	0.485	0.861	0.035	0.673	0.248	0.052	0.44
NDVI	0.286	0.04	0.468	0.218	0.066	0.175	0.088	0.073	0.02	0.201

Table 4.5: *Aedes albopictus* full variable set preliminary model variable importance.

	GLM	GBM	GAM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT
bio1	0.289	0	0.216	0.011	0.09	0.248	0.173	0.006	0.01	0.017
bio2	0.06	0.017	0.048	0.031	0.031	0.249	0.025	0.018	0.082	0.012
bio3	0.279	0.003	0.277	0.13	0	0.193	0.185	0.261	0.016	0.018
bio4	0.528	0	0.352	0.016	0.197	0.242	0	0	0.015	0.008
bio5	0.131	0	0.875	0.015	0.007	0.305	0.363	0.202	0.012	0.014
bio6	0.262	0	1	0	0.334	0.237	0.626	0.136	0.017	0.019
bio7	0.742	0.004	0.607	0.104	0.511	0.256	0.071	0.417	0.05	0.009
bio8	0.001	0.001	0.047	0	0.03	0.275	0	0	0.008	0.038
bio9	0.007	0	0.1	0	0.049	0.249	0	0	0.008	0.038
bio10	0.835	0.01	0.581	0	0.004	0.281	0.619	0.338	0.007	0.017
bio11	0.164	0	0.292	0	0.012	0.236	0.624	0.129	0.011	0.023
bio12	0.041	0.001	0.009	0.057	0.482	0.299	0	0	0.011	0.015
bio13	0.207	0.001	0.274	0	0.295	0.276	0	0	0.01	0
bio14	0.029	0.001	0.04	0.003	0.016	0.229	0.04	0.069	0.012	0.01
bio15	0.007	0	0.02	0.002	0.004	0.138	0	0	0.015	0.203
bio16	0	0	0.051	0.006	0.013	0.278	0	0	0.015	0.002
bio17	0.019	0.002	0.027	0	0.037	0.243	0.046	0.039	0.006	0.022
bio18	0.147	0.474	0.075	0.377	0.315	0.287	0.44	0.122	0.114	0.066
bio19	0.003	0.005	0.011	0	0.012	0.213	0.013	0.008	0.013	0.019
NDVI	0.027	0.009	0.021	0.024	0.008	0.26	0.026	0.032	0.008	0.006

Table 4.6: AUC of ten models for five species (with reduced variable sets). Bolded models were shown in the final models. Updated Zika model incorporating New World outbreak data included as "ZIKV+".

	GLM	GBM	GAM	CTA	ANN	SRE	FDA	MARS	RF	MAXENT
A.Aeg	0.975	0.980	0.981	0.977	0.957	0.855	0.974	0.976	1.000	0.930
A.Afr	0.983	0.999	1.000	0.985	0.967	0.837	0.959	0.979	1.000	0.739
A.Alb*	0.919	0.942	0.938	0.945	0.882	0.760	0.923	0.930	1.000	0.940
ZIKV	0.934	0.975	0.968	0.920	0.773	0.741	0.934	0.938	1.000	0.807
ZIKV+	0.920	0.975	0.946	0.921	–	–	0.927	0.936	1.000	–
DENG*	0.919	0.942	0.938	0.945	0.882	0.760	0.923	0.930	1.000	0.940

Table 4.7: Zika final model variable importances.

	GLM	GBM	GAM	CTA	FDA	MARS	RF
bio1	0.608	0.014	0.654	0	0	0	0.026
bio2	0.919	0.006	0.738	0.068	0	0	0.019
bio3	0.774	0.02	0.481	0.178	0.1	0	0.029
bio4	1	0	0.329	0	0	0	0.017
bio5	0	0.006	0.708	0	0	0.279	0.015
bio6	0.626	0.211	0	0.879	0.46	0.803	0.125
bio7	0	0.033	0.563	0	0.268	0.354	0.051
bio10	1	0.003	0.792	0	0	0	0.014
bio11	0	0.018	1	0	0	0.033	0.047
bio12	0	0.022	0.104	0	0.227	0	0.028
bio13	0	0.109	0.075	0	0.462	0.873	0.116
bio15	0	0.046	0.207	0.323	0.071	0.295	0.026
bio16	0	0.01	0	0	0.153	0.506	0.043
bio17	0.258	0.008	0.183	0.094	0.208	0.211	0.013
bio19	0	0.01	0.02	0.201	0.062	0.326	0.008
NDVI	0	0.082	0.266	0	0.095	0.027	0.06

Table 4.8: Dengue final model variable importances

	GLM	GBM	GAM	CTA	FDA	MARS	RF	MAXENT
bio1	0.388	0.001	0.363	0.004	0.248	0	0.017	0
bio2	0.004	0.029	0.054	0.097	0	0	0.053	0.021
bio4	0.159	0.012	0.241	0.076	0	0.233	0.06	0.031
bio6	0.236	0.008	0.213	0	0.266	0	0.045	0.017
bio7	0.702	0.036	0.571	0.328	0.639	0.406	0.073	0.043
bio8	0.225	0.009	0.083	0.055	0.006	0.235	0.021	0.084
bio9	0.272	0	0.063	0.008	0	0.186	0.012	0.029
bio10	0.427	0	0.502	0.033	0.508	0.051	0.015	0.071
bio11	1	0.109	0.818	0.527	0.85	0.158	0.062	0.487
bio13	0.012	0.008	0.138	0.098	0.072	0.104	0.02	0.001
bio14	0.032	0.009	0.082	0.147	0.022	0.028	0.034	0.03
bio15	0	0.001	0.031	0.028	0	0	0.011	0.03
bio16	0	0	0.075	0.005	0	0	0.016	0.007
bio18	0.005	0.002	0.004	0.011	0	0	0.025	0.001
bio19	0.032	0.026	0.055	0.064	0.069	0.052	0.048	0.074
NDVI	0.095	0.042	0.085	0.17	0.081	0.113	0.05	0.052

Table 4.9: *Aedes aegypti* final model variable importances

	GLM	GBM	GAM	CTA	ANN	FDA	MARS	RF
bio1	0.276	0.01	0.462	0.012	0.019	0.212	0.024	0.02
bio2	0.13	0.011	0.05	0.234	0.023	0	0	0.083
bio3	0.372	0.005	0.148	0.048	0.008	0.102	0.078	0.018
bio4	0.82	0.001	0.198	0.026	0.048	0	0	0.02
bio5	0.273	0	0.178	0.012	0.064	0.215	0	0.016
bio7	0.755	0.019	0.328	0.056	0.057	0.071	0.23	0.048
bio9	0.068	0	0.016	0.039	0.051	0	0.008	0.01
bio10	0.94	0.018	0.611	0.061	0.017	0.716	0.306	0.029
bio11	0.472	0.001	0.621	0.092	0.169	0.752	0.015	0.022
bio13	0.345	0.012	0.06	0.076	0.077	0.753	0.108	0.031
bio14	0.013	0.004	0.023	0.007	0.02	0.019	0.008	0.018
bio15	0.005	0	0.014	0.003	0.032	0	0	0.013
bio16	0.081	0	0.033	0.009	0.088	0.11	0.051	0.024
bio18	0.086	0.14	0.085	0.159	0.661	0.039	0.352	0.065
bio19	0.009	0.004	0.011	0.043	0.026	0.025	0.015	0.018
NDVI	0.051	0.02	0.031	0.034	0.029	0.052	0.042	0.021

Table 4.10: *Aedes africanus* final model variable importances

	GLM	GBM	GAM	CTA	ANN	FDA	MARS	RF
bio1	0.562	0	0.696	0	0.112	0.518	0.062	0.004
bio2	0.184	0.005	0.1	0.078	0.165	0	0	0.007
bio3	0.418	0.005	0.511	0.138	0.01	0.041	0	0.027
bio4	0.999	0.423	0.324	0.717	0.884	0.269	0.548	0.041
bio7	0.41	0.005	0.311	0.101	0.349	0	0.018	0.017
bio8	0	0.013	0.564	0.229	0.451	0.014	0.361	0.012
bio9	0.545	0.002	0.582	0	0.392	0	0.396	0.007
bio10	0.69	0.003	0.584	0.141	0.132	0	0.219	0.009
bio11	0	0.01	0.571	0.276	0.355	0.972	0	0.01
bio13	0.688	0.025	0.747	0.18	0.912	0.871	0.445	0.024
bio14	0.301	0.002	0.265	0.013	0.216	0.071	0	0.004
bio15	0.391	0.001	0.396	0.038	0.293	0	0	0.005
bio16	0.136	0.006	0.398	0.033	0.433	0.356	0.189	0.017
bio18	0.162	0.01	0.331	0.021	0.226	0.069	0.069	0.009
bio19	0.11	0.127	0.517	0.246	0.541	0.733	0.421	0.047
NDVI	0.239	0.065	0.297	0	0.291	0.176	0.21	0.028

Table 4.11: *Aedes albopictus* final model variable importances.

	GLM	GBM	GAM	CTA	FDA	MARS	RF	MAXENT
bio1	0.35	0	0.357	0	0.196	0.198	0.01	0.143
bio2	0.058	0.032	0.001	0.233	0.003	0.018	0.105	0.094
bio3	0.297	0.003	0.069	0.055	0.103	0.189	0.027	0.066
bio4	0.416	0	0.312	0	0.025	0.141	0.023	0.005
bio5	0.569	0	0.41	0	0.286	0.311	0.026	0.019
bio6	1	0	0.45	0	0.041	0.355	0.023	0.001
bio7	0.29	0.006	0.181	0	0	0	0.083	0.003
bio10	0.761	0.011	0.61	0.01	0.646	0.36	0.012	0.122
bio11	0.5	0	0.243	0.007	0.313	0.281	0.013	0
bio12	0.046	0.002	0.017	0	0	0	0.014	0.007
bio13	0.226	0.001	0.055	0	0	0	0.022	0.16
bio14	0.017	0.002	0.012	0	0.008	0.023	0.021	0.028
bio15	0.003	0.001	0.001	0.009	0	0.009	0.016	0.004
bio18	0.105	0.462	0.07	0.643	0.357	0.113	0.154	0.057
bio19	0.002	0.005	0.009	0	0.009	0	0.021	0.021
NDVI	0.026	0.007	0.079	0.002	0.022	0.015	0.009	0.058

Table 4.12: Variable importance in supplementary ZIKV+ model.

	GLM	GBM	GAM	CTA	FDA	MARS	RF
bio1	0.34	0.001	0.431	0	0.125	0.046	0.008
bio2	0	0.006	0.098	0.143	0	0	0.015
bio3	0.223	0.003	0.027	0.231	0.307	0.039	0.024
bio4	0.456	0.094	0.819	0	0.16	0.684	0.07
bio6	0	0.004	0.344	0	0	0	0.011
bio7	0.276	0.025	0	0	0.104	0	0.034
bio8	0.822	0.123	0.231	0.604	0.91	0.501	0.058
bio10	0.636	0.007	0.549	0.198	0	0.59	0.007
bio11	0	0.003	0.994	0	0.036	0.406	0.017
bio13	0	0.009	0.04	0.272	0.082	0	0.018
bio15	0	0.011	0.167	0.017	0.124	0.328	0.017
bio16	0	0.011	0.048	0.061	0.013	0.033	0.012
bio17	0	0.003	0.107	0	0	0.133	0.012
bio18	0.071	0.008	0.1	0.034	0.023	0.163	0.016
bio19	0	0.032	0.14	0.173	0.129	0.111	0.026
NDVI	0.184	0.058	0.061	0.082	0.109	0.156	0.04

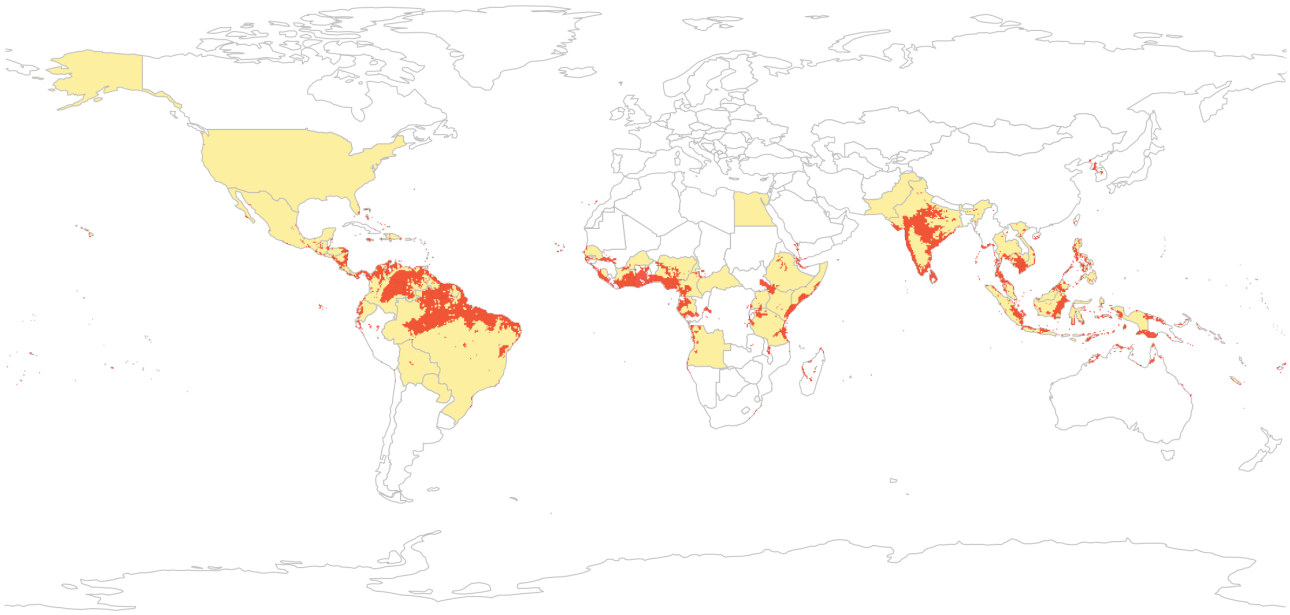


Figure 4.1: The global distribution of case reports of Zika virus (1947 to February 2016) broken down by country (yellow shading) and an ensemble niche model built from occurrence data (red shading). Our model correspond well to shaded countries, with only minor discrepancies (Paraguay, the Central African Republic; a single case in Egypt in the 1950s), We emphasize that displaying cases at country resolution overstates the distribution of the virus, especially in the Americas (for example, Alaska, a point of significant concern given Messina *et al.*'s presentation of their niche model in terms of "highly suitable" countries with broad geographic expanse like the United States, China, and Argentina.

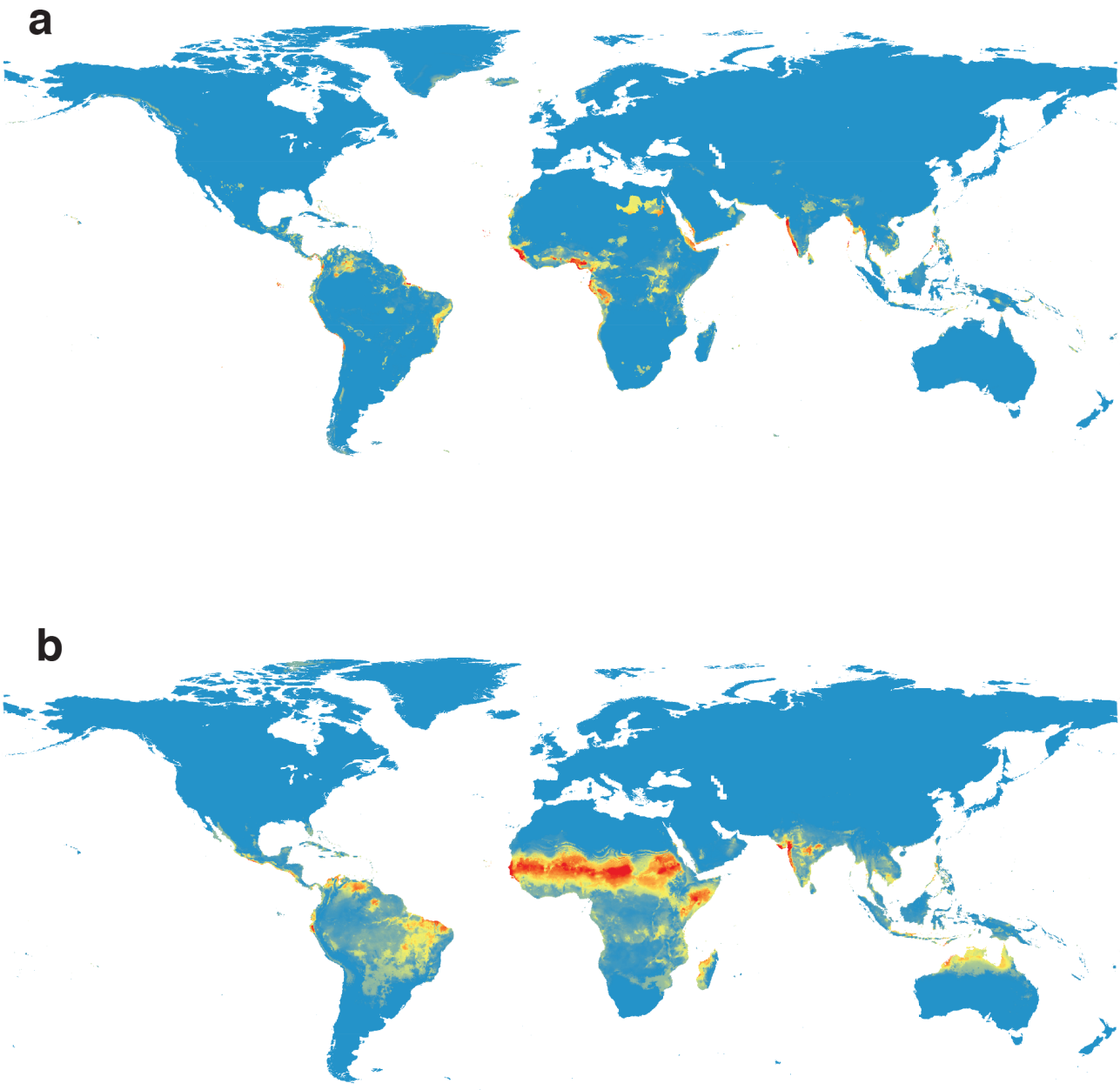


Figure 4.2: Geographical cross validation of (a) the sub-model built from occurrences on the African continent (n = 27) as projected upon the global climate space and (b) the sub-model built from occurrences on the Asian continent (n = 33) projected at the global scale.

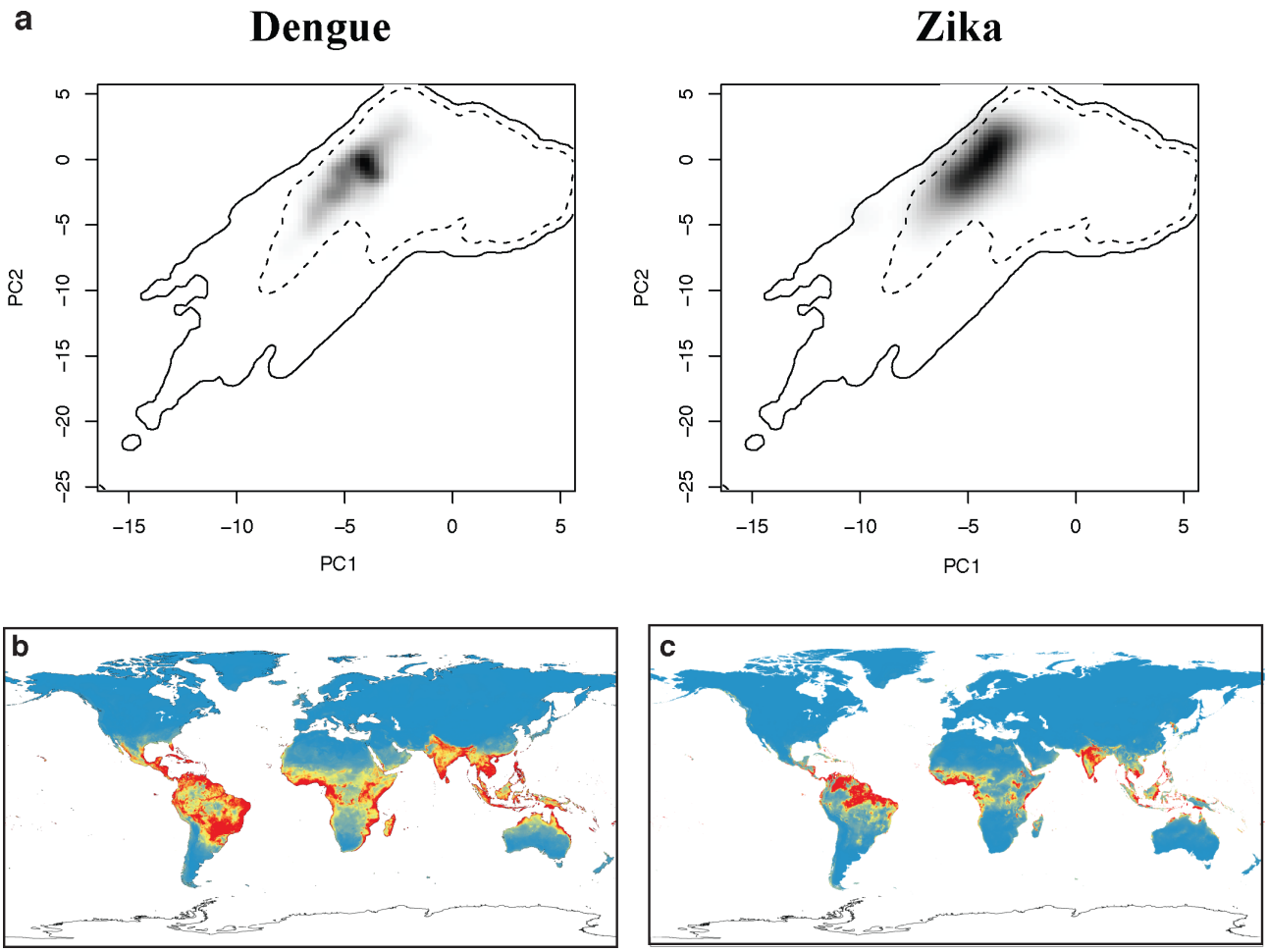


Figure 4.3: The ecological niche of Zika and dengue in principal component space (a). Solid and dashed lines are 100% and 50% boundaries for all environmental data, respectively. Despite apparent overlap in environmental niche space, the dissimilarity between the black shading in each principal component graph indicates statistically significant differences between the niches, evident in the projections of our niche models for dengue (b) and Zika (c).

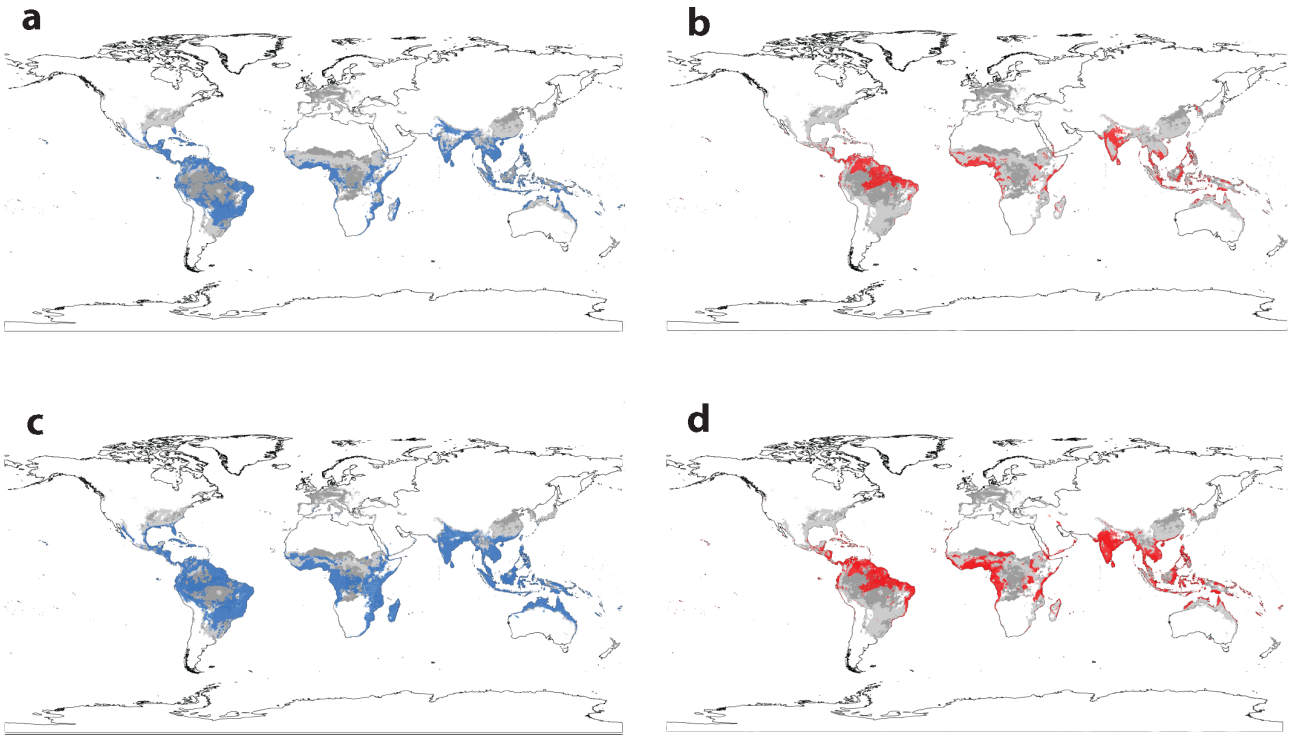


Figure 4.4: The estimated global distribution of Zika (red) and dengue (blue) based on current (a, b) and 2050 climate projections (c, d), compared against the current (light grey) and future distribution (dark grey) of all three mosquito vectors *Aedes aegypti*, *Ae. africanus*, and *Ae. albopictus* (a-d).

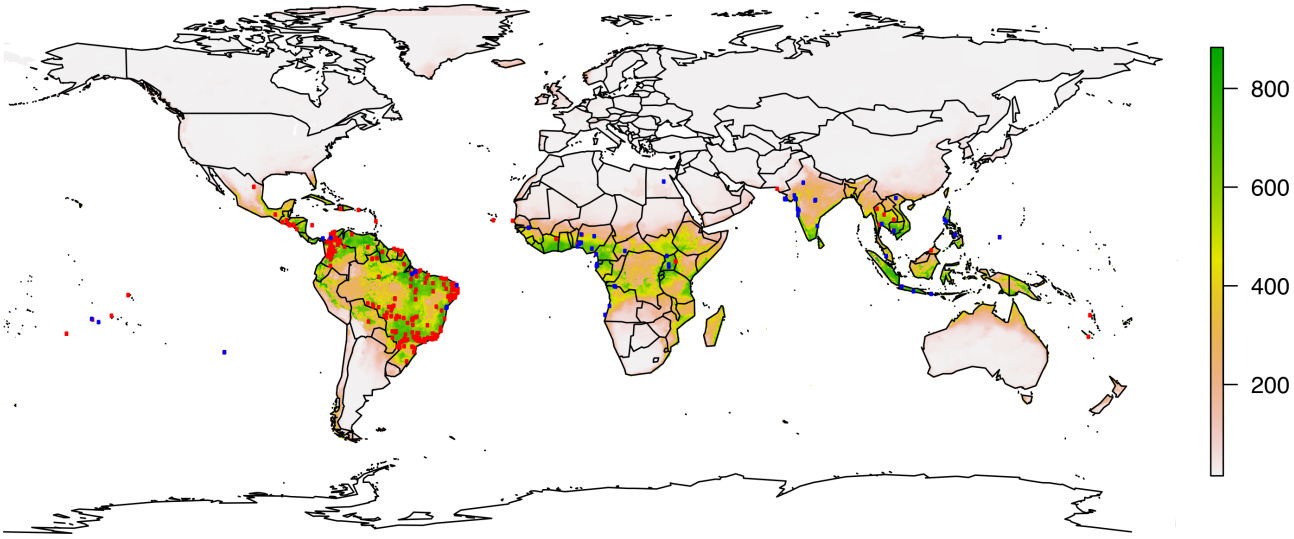


Figure 4.5: An updated ecological niche model incorporating aggregated global data, with Messina *et al.*'s full dataset (red) and ours (blue) against the updated weighted ensemble model.

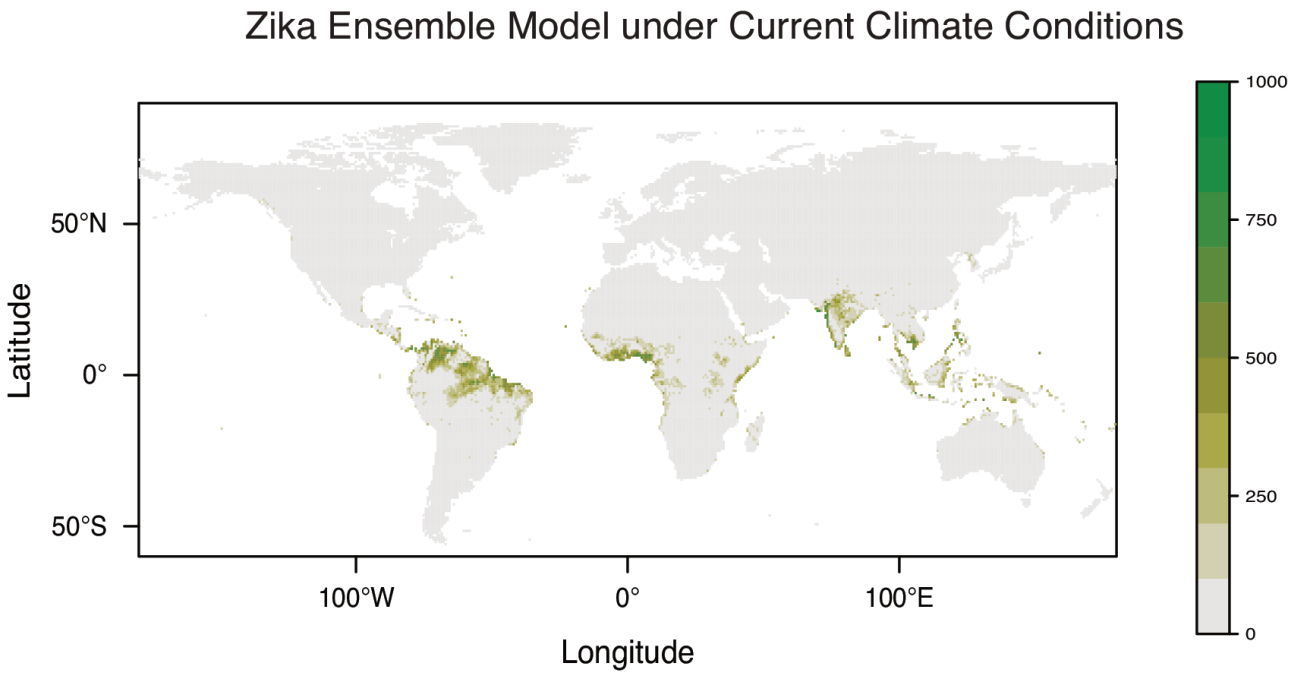


Figure 4.6: Final ensemble model for Zika virus.

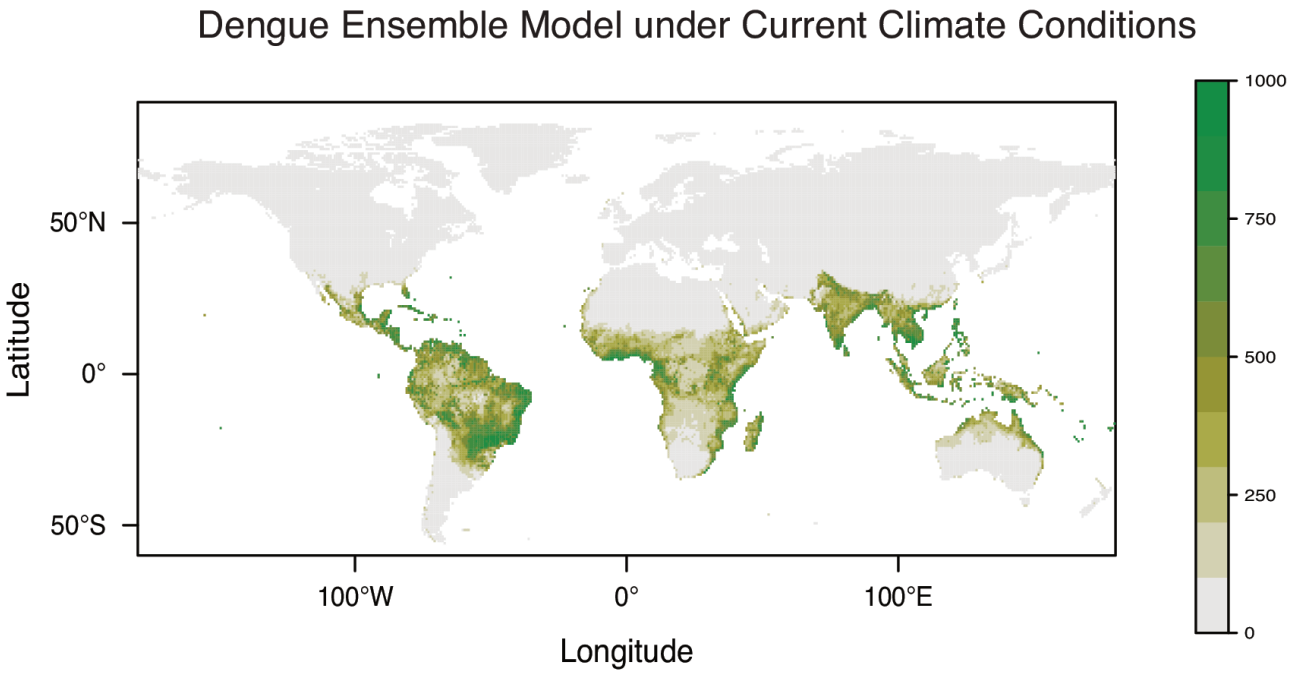


Figure 4.7: Final ensemble model for dengue fever.

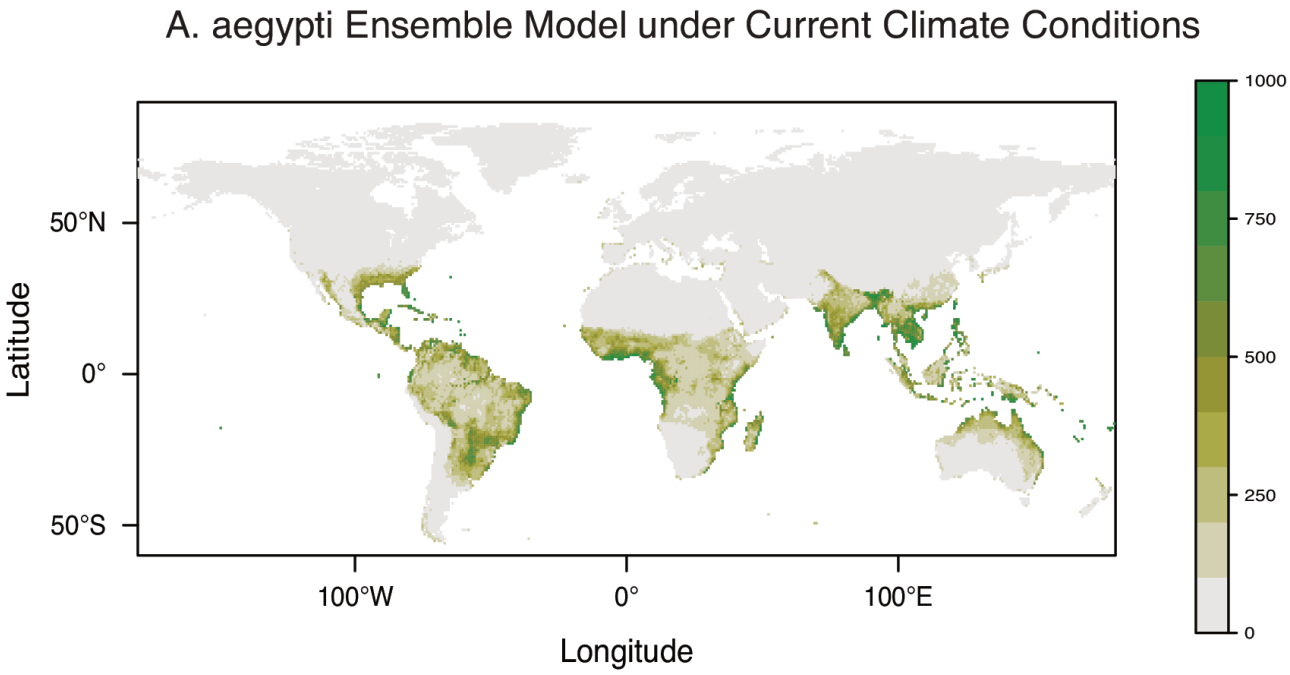


Figure 4.8: Final ensemble model for *Aedes aegypti*.

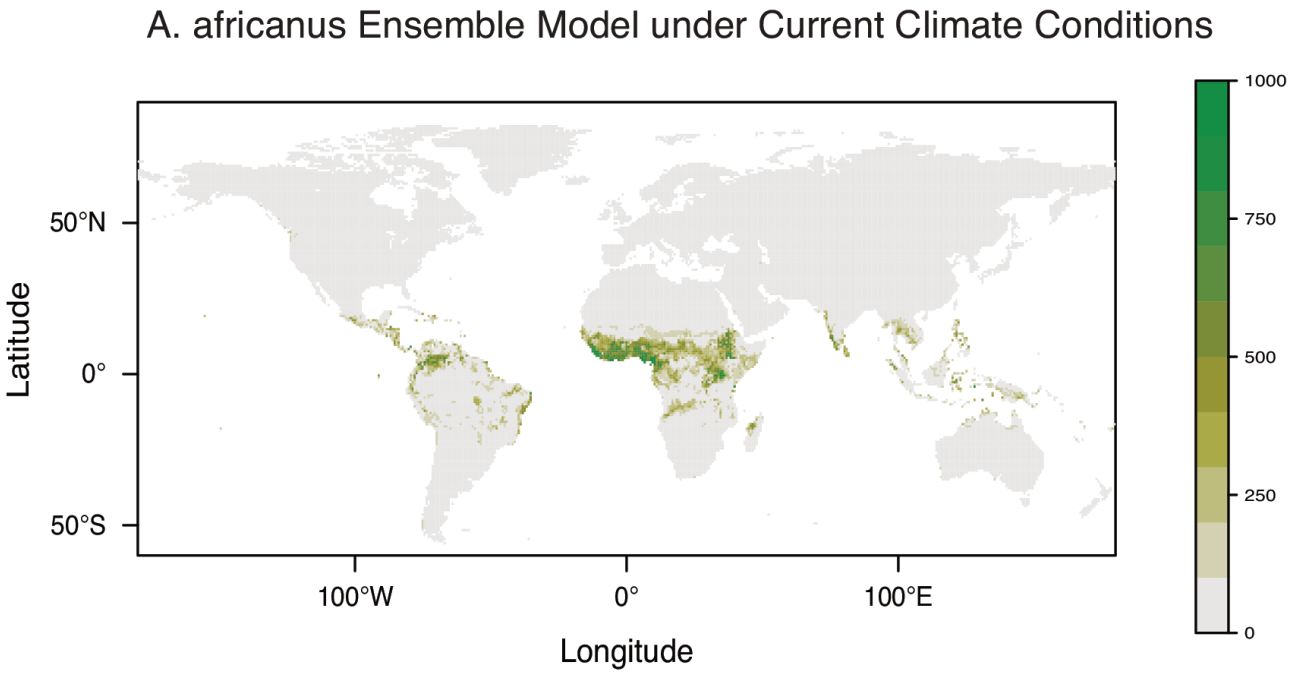


Figure 4.9: Final ensemble model for *Aedes africanus*.

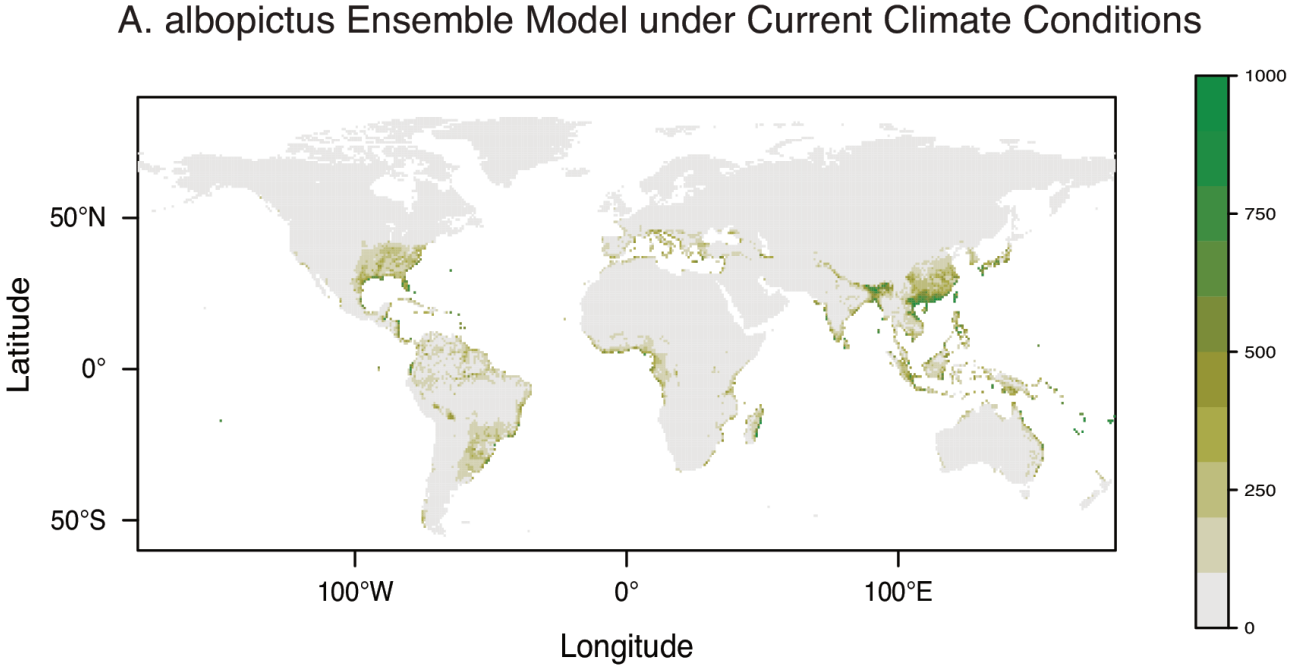


Figure 4.10: Final ensemble model for *Aedes albopictus*.

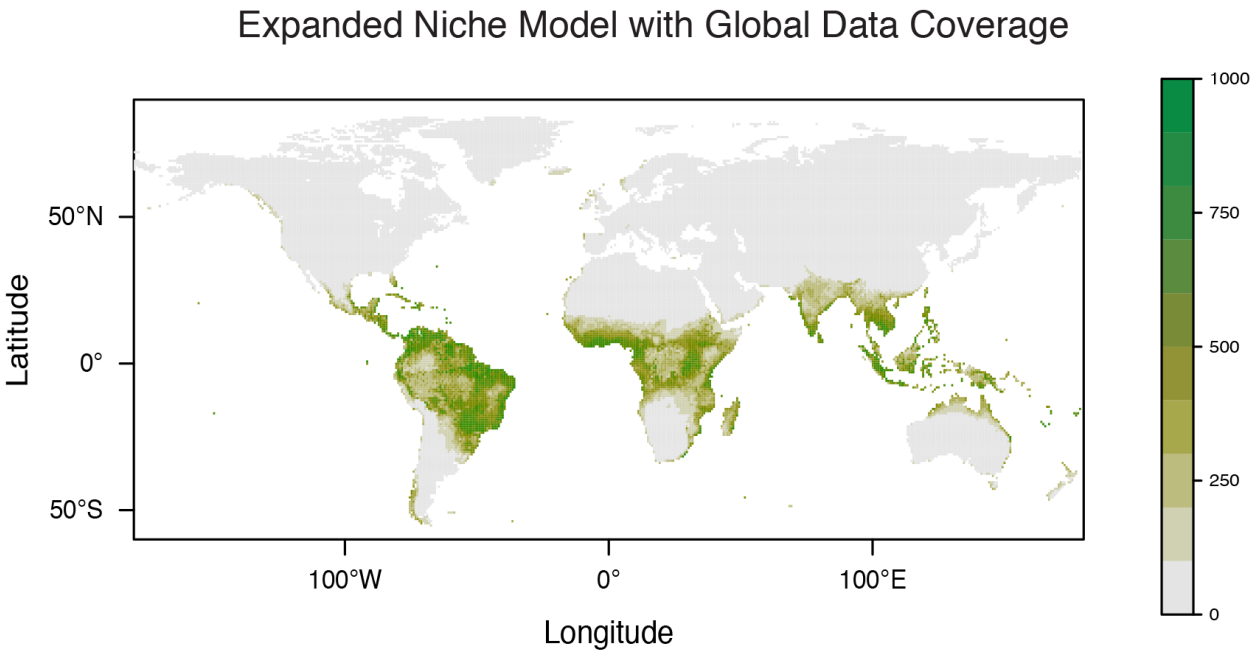


Figure 4.11: Expanded niche model with global data coverage.

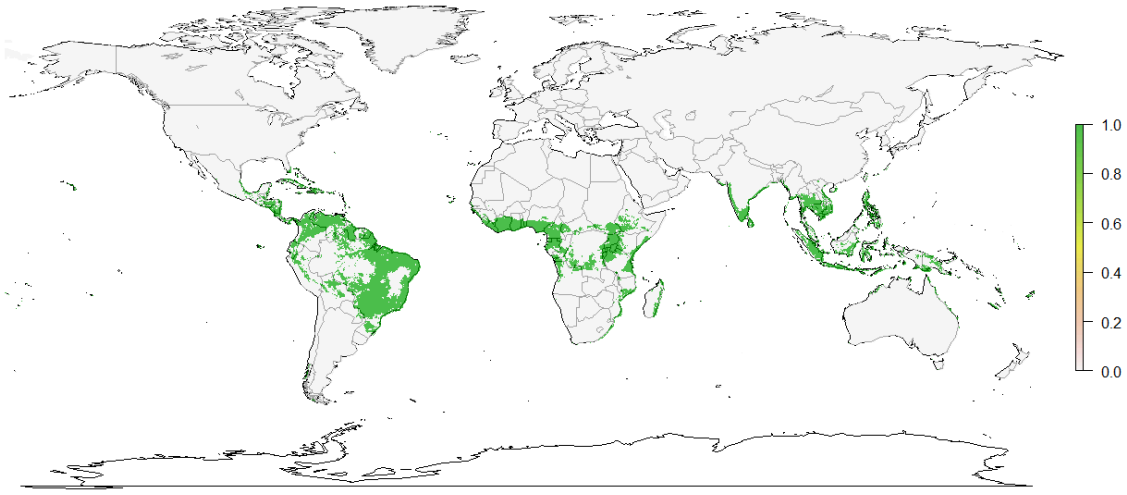


Figure 4.12: Expanded niche model with threshold.

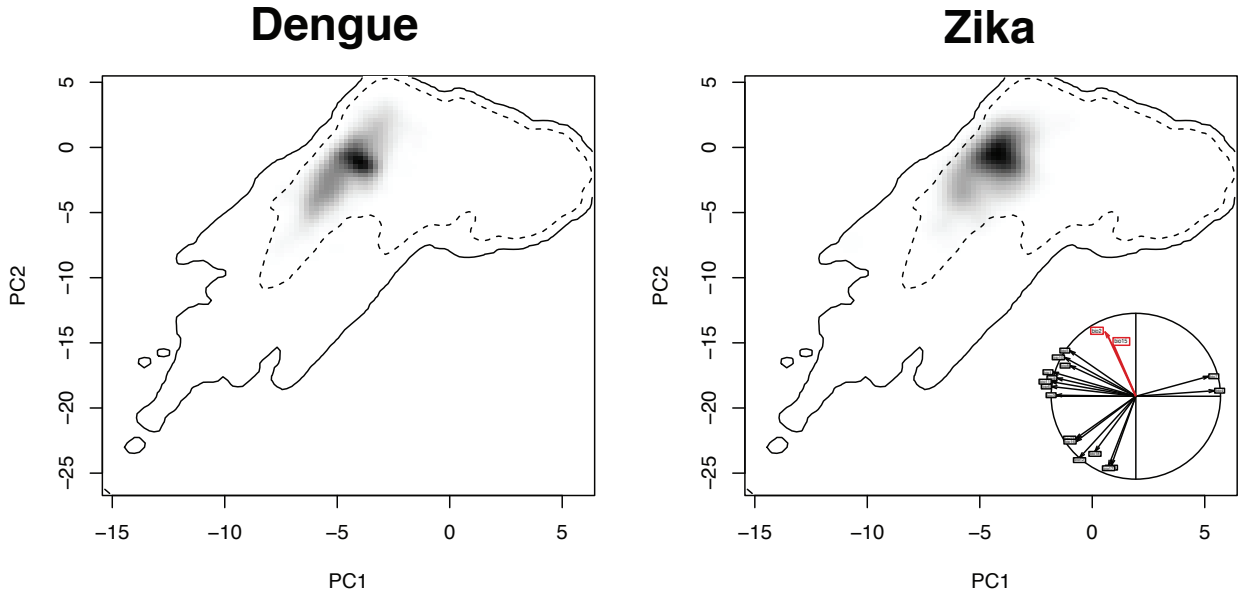


Figure 4.13: Niche overlap analysis between dengue and global Zika database. In the equivalency test, we find significant evidence for differences (Schoener's D = 0.295; p = 0.004).

Chapter 5

Consensus and conflict among ecological forecasts of Zika virus outbreaks in the United States

Colin J. Carlson Eric Dougherty Mike Boots Wayne Getz Sadie J. Ryan

5.1 Abstract

Ecologists are increasingly involved in the pandemic prediction process. In the course of the Zika outbreak in the Americas, several ecological models were developed to forecast the potential global distribution of the disease. Conflicting results produced by alternative methods are unresolved, hindering the development of appropriate public health forecasts. We compare ecological niche models and experimentally-driven mechanistic forecasts for Zika transmission in the continental United States, a region of high model conflict. We use generic and uninformed stochastic county-level simulations to demonstrate the downstream epidemiological consequences of conflict among ecological models, and show how assumptions and parameterization in the ecological and epidemiological models propagate uncertainty and produce downstream model conflict. We conclude by proposing a basic consensus method that could resolve conflicting models of potential outbreak geography and seasonality. Our results illustrate the unacceptable and often undocumented margin of uncertainty that could emerge from using any one of these predictions without reservation or qualification. In the short term, ecologists face the task of developing better post hoc consensus that accurately forecasts spatial patterns of Zika virus outbreaks. Ultimately, methods are needed that bridge the gap between ecological and epidemiological approaches to predicting transmission and realistically capture both outbreak size and geography.

5.2 Introduction

In the urgent setting of pandemic response, ecologists have begun to play an increasingly important role. [283] Ecological variables like temperature and precipitation often play just as important a role as socioeconomic risk factors in the vector-borne transmission cycle, governing key parameters including transmission rates, vector lifespan, and extrinsic incubation period [284]; the statistical relationships among these variables can be exploited to develop predictive frameworks for vector-borne disease outbreaks. These models are often developed for mechanistic prediction at local scales, but ecologists have recently begun to play a more important role in predicting the overall possible distribution of emerging infections. Ecological niche modeling is a typically phenomenological method that correlates occurrence data with environmental variables to make inferences about the geographic boundaries of potential transmission. [285] Within niche modeling approaches, there are conflicting views regarding which algorithms are appropriate to use in the context of particular applications [167, 286, 109], and consensus methods have hardly advanced beyond basic model averaging. [287] As an increasingly popular alternative, mechanistic ecological models have been developed that extrapolate geographic projections from experimental results [288], but these can be data-intensive and highly sensitive to parameterization. [289, 290, 288] In theory, the two approaches—phenomenological and top-down, or mechanistic and bottom-up—should be roughly congruent when implemented with sufficient data and predictors, as they approximate the same pattern. [291] Yet discrepancies between the two approaches in practice highlight a tension in species distribution modeling, between deductive approaches that infer ecology from observed broad-scale patterns, and inductive approaches that scale ecological experiments to predict real patterns. In the context of pandemic response, the trade-off has acute stakes: early access to ecological predictions can help pandemic efforts, but inaccurate information based on limited data could drive misallocation of public health resources. [292]. Thus, there is a clear need to develop better consensus methods, but even before that, a need to understand the epidemiological implications of the differences among model-building approaches.

The Zika virus (henceforth Zika) pandemic that was first detected in Brazil in 2015 highlights the unusual and sensitive challenges of pandemic response. A number of characteristics make Zika unique from a public health standpoint, including its rapid spread through the Americas after a slow, multi-decade spread from Africa through Asia; the appearance of a sexual route of transmission, a rare feature for a vector-borne pathogen; and perhaps most importantly, the appearance of high rates of microcephaly, and more broadly the emergence of Zika congenital syndrome. At least 11,000 confirmed cases of Zika have affected pregnant women, leading to roughly 10,000 cases of birth defects, including microcephaly. [293] As of April 6, 2017, a total of 207,557 confirmed cases of autochthonous transmission (out of 762,036 including suspected cases) have been recorded in the Americas. [294] Moreover, Zika is exceptional among vector-borne diseases in that it has developed a sexual pathway of transmission in humans (comparable examples, such as canine leishmaniasis, are incredibly rare [295]). The rapid spread of Zika virus from Brazil throughout the Americas has posed

a particular problem for ecologists involved in pandemic response, as several different ecological niche models (ENMs) [292, 12, 296] and a handful of mechanistic forecasts [290] have been developed to project the potential full spread of the pathogen. So far, autochthonous transmission has been recorded throughout most of Central America and the Caribbean, with cases as far north as the southern tips of Texas and Florida.

In this study, we focus on the United States as a test system for exploring conflict between different model predictions. A Brazil-scale outbreak of Zika in the United States could be devastating; one model for only six states (AL, FL, GA, LA, MS, TX) found that even with the lowest simulated attack rate, Zika outbreaks could be expected to cost the United States over $180 million, and estimates under worse scenarios exceed $1 billion. [297] Consequently, a high priority has been placed on developing accurate models that capture socioecological suitability for Zika outbreaks in the United States. [298, 299] However, we suggest that the lack of a consensus among different models of spatial risk renders the literature less credible or navigable to policymakers, as predictions under certain conditions span a range from 13 counties at risk [292] to the entire United States (Figure 5.1). [290, 300] At the time of writing, the majority of public health agencies in the United States were preparing for the apparent eventuality of Zika, based either on no prior geographic information, or basic data on the range of *Aedes* mosquitoes. [301, 302] Millions of dollars have already been invested in state- and city-level Zika preparation, even in areas without recently-recorded *Aedes* presence, and pesticide spraying for vector control has already had unanticipated consequences, including killing millions of honeybees. [303] Domestic efforts to prepare for Zika are not unreasonable in the absence of a consensus prediction about Zika's likely final range; the continued importation of new cases into every state in the U.S. likely amplifies the perceived threat of local outbreaks, especially given the pathway of sexual transmission (which could conceivably start stuttering chains [304] outside regions of vector-borne transmission). However, an informed response to Zika in the United States requires both a greater consensus about at-risk areas, and a more precise understanding of the uncertainty contained in different ecological forecasts.

Ecological niche models for vectors and pathogens are commonly used as an underlying foundation in epidemiological models, or more broadly, spatial studies in public health and policy work (including in the Zika literature [305, 306, 307, 298]) In this study, we highlight the unavoidable – but usually, unacknowledged – downstream consequences of model selection in those cases, and illustrate the lack of any one clear way to resolve conflict among published, peer-reviewed ecological studies. To expose this problem more clearly, we compare four published ecological predictions for the extent and duration of possible Zika virus transmission in the United States, and overlay generic epidemiological forecasts to measure the impacts of model differences. In doing so, we examine the scale of epidemiological uncertainty introduced at five scales:

1. Different environmental variable selections for a given niche modeling approach [296]

2. Differences among published ecological niche models

3. Differences between phenomenological [292, 12, 296] and mechanistic [290] approaches
4. Differences driven by parameterization of Bayesian mechanistic models [290]
5. Differences in how population-at-risk is aggregated from the niche models for epidemiological simulations

In the process, our exercise shows that relying on any one ecological model adds a hidden layer of uncertainty to epidemiological forecasts, indicating the need to develop better consensus methods—and to develop ecological and epidemiological tools in a more integrated approach that better approximates observed outbreak patterns.

5.3 Methods

Ecological Models

Three studies have been published using ecological niche models (ENMs) to map the possible distribution of Zika virus, using a different combination of occurrence data, environmental predictors, and statistical approaches [292, 12, 296]. Their models suggest varying degrees of severity, especially as measured within the United States (Table 5.1). Other models have also been widely used in epidemiological work as a proxy for the distribution of Zika, such as an ecological niche model of *Aedes aegypti* and *Ae. albopictus*[253] (fairly commonly used, e.g., [308, 309]; or see [310], which presents its own *Aedes* ENM that becomes a risk map of Zika transmission), or dengue-specific niche models (recently used by Bogoch *et al.* in two separate publications [270, 305]). Most ecological niche models indicate the range of Zika virus should be more restricted than that of its vectors, and published evidence suggests there may be significant differences between the known and potential distributions of dengue and Zika [292], so we exclude these proxy methods from our study and focus instead on modeling studies that explicitly use Zika occurrence data.

Carlson *et al.*

Carlson *et al.*[292] developed an ensemble niche model constructed using the R package BIOMOD2. The resulting model uses seven of ten possible methods (general linear models, general additive models, classification tree analysis, flexible discriminant analysis, multiple adaptive regression splines, random forests, and boosted regression trees), notably omitting maximum entropy (MaxEnt). Their primary model uses only occurrence data from outside the Americas, but here we adapt their secondary model which incorporates data from Messina *et al.* (below) to show the lack of the sensitivity of the method to that additional data, especially in the United States. The only environmental predictors used are the BIOCLIM dataset [171] and a vegetation index (NDVI). The final model threshold was selected to maximize the true skill statistic, with a selected value of 0.271 used in the original study to produce a binary suitability map. In the Carlson *et al.* model, suitable range for Zika virus

is predicted to be limited to the southern tip of Florida and small patches of Los Angeles and the San Francisco Bay area. Only a total of 13 counties have any suitable area in this model; at the county scale, this model has the greatest concordance with observed outbreak patterns of autochthonous transmission in the United States during 2016.

Messina et al.

Messina *et al.*[12] use an ensemble boosted regression trees approach with a global dataset of occurrence points primarily from South American outbreak data. The model incorporates prior information about *Aedes* distributions. For example, pseudoabsences are preferentially generated in areas of lower *Aedes* suitability. Their model uses six environmental predictors: two direct climate variables, two indices of dengue transmission based on temperature (one for *Ae. albopictus* and one for *Ae. aegypti*), a vegetation index (EVI), and a binary land cover classifier (urban or rural). Messina *et al.* select a threshold of 0.397 that marks 90% of occurrence data as suitable (10% omission). Their model predicts that suitable range for Zika virus encompasses a substantial portion of the Gulf Coast, including the entirety of Florida and as far west as eastern Texas. Their study is also the first to estimate population-at-risk, placing the global figure at 2.17 billion people.

Samy et al.

Samy *et al.*[296] use MaxEnt to build four sub-models with different combinations of environmental predictors. The first is a conventional ENM approach using environmental predictors (precipitation, temperature, EVI, soil water stress, "aridity," and elevation). The second, a more unconventional approach in the niche modeling literature, separates out socioeconomic predictors (among them population density, night light from satellite imagery, and a function of expected travel time called "accessibility"). The third uses all the same as the first model but with three added layers (land cover and suitability for *Ae. aegypti* and *Ae. albopictus*); finally, in model 4, all variables are included and we use that here as the representative case of the alternative Samy formulations. (In an additional sub-analysis, we compare these four models and show the impacts of these variable selection choices on downstream epidemiological forecasts.) For all, the model threshold is selected based on a maximum 5% omission rate for presence data, and also projects high environmental suitability in the Gulf region, very similar to that of Messina *et al.*. This model also produced isolated suitable patches based on social factors, which predominantly occur at urban centers.

Mordecai et al.

Mordecai *et al.*[290] produced a Bayesian model of transmission of *Aedes*-borne viruses (dengue, chikungunya, and Zika) in the Americas that we adapt as a mechanistic geographic forecast for subsequent analyses. In their main model, an R_0 modeling framework is constructed based on models for vector borne diseases, building upon the Kermack-McKendrick R_0 model for malaria [311]. In this model, the majority of parameters describing the life

cycle of mosquitoes and parasite development within the mosquitoes are sensitive to temperature. Mordecai *et al.* used data derived from the literature to parameterize the shape of the temperature response for each temperature sensitive parameter. These are based on laboratory observations of *Aedes aegypti* and *Aedes albopictus*, and infections with dengue, chikungunya, and Zika at constant temperatures through the range of possible values. Because these are bioenergetic functions, curve fitting exercises to derive appropriate models of the non-linear relationships underlie the parameterization of the overall transmission model. A non-linear overall relationship between transmission (R_0) and temperature is fitted in a Bayesian inference framework, and from it two endpoints of a "suitable range" can be extrapolated within which $R_0 > 0$. Those ranges can be adjusted for different levels of posterior probability, and can be used as a suitability threshold that can be projected onto gridded temperature data, producing binary monthly maps of suitability (which can be aggregated to year-round possible presence). In the Mordecai *et al.* publication, the most conservative probability level ($> 97.5\%$) was then mapped onto long-term mean monthly average temperatures in the Americas, derived from Worldclim data [171], to estimate the number of months transmission was possible for *Ae. aegypti* and *Ae. albopictus*[290]. Additional maps were also constructed of the number of months of possible transmission for $R_0 > 0$ at posterior probabilities of 50% and 2.5%, and are found in the supplemental material. Here, we use all three probability levels from the *Ae. aegypti* model, to project the terms of the number of months of predicted transmission potential by mapping the model onto WorldClim temperature gridded data for long-term monthly minimum and maximum temperatures (six possible combinations).

Consensus Mapping Methods

In a preliminary effort to present a consensus forecast based on current ecological understanding, we use two alternative methods to develop county-scale predictions from the models included in our analysis. The first ("majority rule") excludes the Mordecai model, and simply applies a majority rule to the binary thresholded Carlson, Messina, and Samy county shapefiles (i.e., any county with agreement between a majority of the niche models for either presence or absence). In the second model ("seasonal majority rule"), we take the counties predicted by the majority rule method and restrict their suitability to the months predicted in the strictest Mordecai model (97.5% confidence) for minimum temperatures. That process excludes 13 of the counties deemed suitable according to the simple majority rule, but which are predicted to be unsuitable year-round in the Mordecai model.

Epidemiological Model

To simulate potential Zika outbreaks in the United States, we adopt the modeling framework used by Gao *et al.*, which incorporates both sexual and vector-borne transmission [312]. We selected Gao *et al.*'s framework because, while fairly simple, it includes a number of important features of the epidemiology of Zika, including the high rate of asymptomatic cases, and

lingering (primarily sexual) transmission by post-symptom "convalescent" cases. Because the transmission term is normalized by dividing by total population size, the model itself is scale-free. Thus, the values associated with each compartment could be represented as proportions rather than the number of individuals. The model divides the human population into six compartments with levels: susceptible (S), exposed (E), symptomatically (I_1) or asymptomatically (A) infected, convalescent (I_2), and recovered (R), where he h and v refer to the human host and mosquito vector populations, respectively:

$$\frac{dS_h}{dt} = -ab\frac{I_v}{N_h}S_h - \beta\frac{\kappa E_h + I_{h1} + \tau I_h 2}{N_h}S_h \tag{5.1}$$

$$\frac{dE_h}{dt} = \theta(ab\frac{I_v}{N_h}S_h + \beta\frac{\kappa E_h + I_{h1} + \tau I_h 2}{N_h}S_h) - \nu_h E_h \tag{5.2}$$

$$\frac{dI_{h1}}{dt} = \nu_h E_h - \gamma_{h1} I_{h1} \tag{5.3}$$

$$\frac{dI_{h2}}{dt} = \gamma_{h1} I_{h1} - \gamma_{h2} I_{h2} \tag{5.4}$$

$$\frac{dA_h}{dt} = (1-\theta)(ab\frac{I_v}{N_h}S_h + \beta\frac{\kappa E_h + I_{h1} + \tau I_h 2}{N_h}S_h) - \gamma_h A_h \tag{5.5}$$

$$\frac{dR_h}{dt} = \gamma_{h2} I_{h2} + \gamma_h A_h \tag{5.6}$$

$$N_h = S_h + E_h + I_{h1} + I_{h2} + A_h + R_h \tag{5.7}$$

Vectors are governed by a complimentary set of equations but only divided into susceptible, exposed, and infected classes:

$$\frac{dS_v}{dt} = \mu_v N_v - ac\frac{\eta E_h + I_{h1}}{N_h}S_v - \mu_v S_v \tag{5.8}$$

$$\frac{dE_v}{dt} = ac\frac{\eta E_h + I_{h1}}{N_h}S_v - (\nu_v + \mu_v)E_v \tag{5.9}$$

$$\frac{dI_v}{dt} = nu_v E_v - mu_v I_v \tag{5.10}$$

$$N_v = S_v + E_v + I_v \tag{5.11}$$

Rather than use the fitted parameters from any given country's outbreak, parameters for the above models were randomly generated from a set of uniform prior distributions specified by Gao *et al.* as reasonable priors based on the literature (Table 5.5). Evidently, these models are significantly discrepant with outbreaks in the continental U.S. so far, with fewer than

300 cases of local transmission recorded in 2016 (and in fact, our simulations are far more severe in terms of final case burden than estimates for Brazil or Colombia). However, the purpose of applying this epidemiological model across the spatial extent predicted by each niche model is both to illustrate the uncertainty that goes unstated in presenting such ENMs and to intimate the necessity of developing and parameterizing these models in concert.

County-Level Simulations

In our main models, every spatial projection of Zika risk was summarized at the U.S. county scale, such that if a single pixel within a county polygon was projected to be suitable under a given model, the county was marked suitable for outbreaks. This assumption clearly overestimates population at risk, but environmental suitability is often aggregated to the county scale in order to develop Zika models for the U.S. [309, 298] For the Mordecai models, the maximum value (months suitable per year) of all pixels within a county was assigned as the value. For example, if a single 25 km^2 cell in a particular county was suitable for a single month, simulations were run for one month with mosquitoes present and the remaining 11 with a mosquito population of zero. While this approach has the potential to overestimate populations vulnerable to mosquito-borne transmission, it adds a number of key strengths. Aggregating information at the county scale absorbs some of the relative spatial uncertainty of predictions at the pixel scale, and may account for source-sink dynamics for vector-borne outbreaks driven by heterogeneity in vector density and competence. Moreover, the county level is one of the finest scales at which public health infrastructure is likely to decide whether interventions like vector control are necessary. Finally, sexual transmission can spread from cells with suitable vectors to vector-free areas, and as a function of both sexual transmission and underlying mobility, outbreaks are therefore unlikely to be contained to a given pixel. Previous work has similarly used the county scale to study risk factors and model outbreak risk for Zika [309, 298], and we follow their precedent.

Population data for each county was taken from projections to the year 2016 based on the 2010 United States Census, and were set as the total susceptible human population at the start of a year. The mosquito population was set at five times the baseline human population, the middle of the range selected by Gao *et al.*. [312] While other studies have used a lower ratio [299], we set mosquito populations (the only parameter we explicitly fixed from the Gao model) as high as we did because many simulations with lower mosquito populations faded out immediately, and setting a higher ratio made the impacts of model differences more immediately apparent. Outbreaks were simulated stochastically at the county level using the Gao *et al.* model, initiated with a single infected person per county. We randomly selected a value for each of the parameters in the Gao *et al.* model for which a range was provided, using a uniform distribution (Table 5.5). For each modeling model, 100 simulations were run in each county designated as suitable. For the three ENMs, county models were run for a "model year" (twelve months of thirty days each), and had no interactive effect on each other. For the Mordecai models, the full vector- and sexually-transmitted epidemic models were run for the number of months (thirty days each) that were predicted suitable. After

that period, the total vector population N_v was set to 0, effectively ending vector-borne transmission, but models continued so that sexual transmission was ongoing up to 360 days. All simulations were run in R 3.3.2, and all scripts and county simulation data are available as supplementary files.

Within-County Heterogeneity

In a final set of analyses, we examine the impact of how risk is aggregated at the county scale. Fine-scale population data does exist for the world from multiple sources[313], but at the resolution niche models are often generated, clear problems exist. Running models on a pixel-by-pixel basis would likely be computationally prohibitive in many cases (including this one); moreover, in the context of sexual transmission, models that do not explicitly include human movement between nearby pixels might produce results that make little sense. While vector movement may be fairly minimal, human movement likely produces mixing at broader geographic scales for both vector-borne and sexual transmission. Aggregating niche models to a county-level suitability is one solution to the problem, and has the added benefit of plausibly absorbing some of the uncertainty among different ENMs. However, this also has the clear tendency to overestimate population at risk; to examine how strongly this affects models, we include an additional set for Carlson, Messina, and Samy (model 4) where susceptible population is scaled down linearly by the proportion of the county marked suitable in each model. This, in itself, adds another layer of neutral assumptions (populations are treated as having a uniform distribution within counties) but might also produce less drastic differences between outbreak trajectories. The results of that analysis are given in Table 5.2 and Figures 5.2-5.4.

5.4 Results

Ecological forecasts for Zika suitability span the range of thirteen counties to all 3108 counties in the continental United States (Table 5.1), and this uncertainty (unsurprisingly) produces tremendous downstream variation in outbreak size. For ENM-based projections, the margin of error among mean trajectories spans an order of magnitude, with a total difference of 168 million cases between Carlson and Samy (Figure 5.1). Areas predicted by other methods to be at the greatest risk from Zika virus are roughly agreed upon among the models, with southern California and the Gulf Coast represented most significantly as outbreak hotspots among the three models. Agreement among all three models is limited, but is most significantly clustered in these areas, especially in the southern tip of Florida and Los Angeles County. While we assumed that aggregating risk at the county level could potentially absorb some of the spatial uncertainty of models and decrease differences between them, we found that it actually substantially exacerbated the observed differences among them (Table 5.2). This was perhaps most notable in the most populous counties, such as Los Angeles county (see supplementary Figures 5.2-5.4).

Model parameterization has a considerable impact on downstream epidemiological results. The four models proposed by Samy, each with slightly different environmental variable selection (see the Methods), produced correspondingly different results (Table 5.3, Figure 5.5). Perhaps counter to our *a priori* expectations, adding more predictors produced broader projections and larger epidemics (not tighter-fit models); the model with all predictors (model 4) produced the largest epidemic, while the one with only climatic covariates (model 1) was in fact somewhat smaller than the Messina outbreak simulation. Model 2 (only social predictors) was only slightly more severe than model 4 (all predictors, which we use as the "Samy model" in all other cases). Adding more predictors increased the projected impacts most noticeably in Arizona and New Mexico; the projections in model 1, the most conservative, were very similar geographically to the Messina model, but with much more substantial range in the Pacific northwest.

A roughly comparable range of predictions to the span of the Carlson, Messina, and Samy models is contained within the entire span of possible implementations of Mordecai's Bayesian model (Figure 5.6). Whereas ENM approaches indicate a somewhat restricted geographic range for possible outbreaks, the Mordecai model suggests that even in a conservative scenario (using minimum temperatures, and 97.5% posterior probability), the majority of *Aedes aegypti*'s range is at least seasonally suitable for Zika transmission. A far greater range of variation is contained within the minimum-temperature-based model scenario, which encompasses roughly half of the land area of the continental U.S. In contrast, the three scenarios based on maximum temperatures are geographically indistinguishable, though worsening projections do extend the seasonality of transmission and thereby produce somewhat longer-tailed epidemics (Figure 5.6a).

Among nine ecological scenarios considered (three niche models and six mechanistic scenarios), an overwhelming spread of possible epidemics could be predicted for the United States (Figure 5.7). The accompanying spatial pattern of case burden also varies between interpretations; while the spatial patterns are roughly identical for Carlson, Messina, and Samy, the temporal dimension introduced by mapping the Mordecai model onto monthly temperature grids dramatically affects how cases are ultimately distributed—and produces a reduction in epidemic size in some scenarios (Figure 5.8). In fact, the most conservative Mordecai scenario (97.5% confidence with minimum temperature) falls between Carlson and Messina in terms of case burden, despite predicting more than four times as many counties with transmission suitability as the Messina model. Across all models, forecasts predict that the majority of the case burden will still be seen along the Gulf Coast and in southern California. The advantage of interfacing ecology and epidemiology is especially evident here; for example, while the original Carlson *et al.* study noted the most significant suitable area was in southern Florida and failed to comment on the potential importance of southern California, the most significant epidemic predicted by most models is in Los Angeles county. The exception is the most conservative Mordecai scenario (Figure 5.6b), the only parameterization of that model in which Los Angeles is designated unsuitable for transmission—a fairly important discrepancy, given that the county is the most populous in the United States, and correspondingly contributes substantially to epidemic size in every other scenario (Figure

5.8a-c).

In an effort to illustrate a method of resolving these conflicting predictions, we present a final "consensus model" that incorporates all four modeling studies. Consensus methods are limited for ecological niche models [287], so we adopt one possible approach: a majority rule at the county scale across Carlson, Messina, and Samy (i.e., in Figure 5.1b, any county value at or above 2 is "suitable," and any below is "unsuitable"). Building on this "majority rule model," for counties that are marked suitable by the ecological niche models, we superimpose the monthly transmission values from Mordecai's most conservative scenario, which most closely matches the geographic extent predicted by the ecological niche models (Figure 5.6a versus Figure 5.8a-c). This filtered "seasonal majority rule" algorithm incorporates the temporal dimension of transmission that is added by our implementation of the Mordecai model while maintaining consensus among the niche models.

The seasonal majority rule model produces a somewhat unsurprising pattern where year-round transmission is most common in the tropics, with seasonal transmission most important in the southeast United States, southeast Brazil, southeast China, and the Himalayas. Unsurprisingly, this produces a comparatively conservative outbreak prediction (Figure 5.10). The inclusion of the temporal component from the mechanistic model reduces case burden by almost two-thirds (Table 5.2), and excludes a handful of counties in the process (which were suitable in the ENM approach but not suitable for a single month in the mechanistic model). Most notably, Los Angeles county (which is suitable for no months of the year in the conservative Mordecai model) is excluded despite being suitable in all three ENMs, which contributes substantially to the overall reduction of projected case totals in the seasonal majority rule approach.

5.5 Discussion

Clear Problems, No Easy Answers

By constructing epidemiological simulations on top of ecological niche models, we found that subtle differences among—and within—rigorous modeling frameworks can introduce substantial downstream variability in outbreak sizes and durations. Conflict among different published studies is the most immediately apparent problem, especially given the lack of a more sophisticated method of resolving these differences beyond the majority rule approach we use. However, for any given study, we show that internal model assumptions carry a level of uncertainty that is hard to understand just from looking at a "final model," a fact that is readily apparent by comparing the different Samy models and Mordecai model parameterizations. (This can be a problem even in cases where no conflict exists among different published models; for example, one model of *Aedes aegypti* and *Ae. albopictus* is most commonly used across purposes [253], including frequently as an outer bound in epidemiological models.) Subjective model design issues like occurrence data collection and thinning, environmental variable selection, pseudoabsence generation, model algorithms, and

threshold selection all introduce subjectivity into niche modeling that goes beyond basic issues of accuracy and exposes deeper strategic tensions in modeling (e.g., Levins' proposed tradeoff in modeling among realism, precision, and generality [314]). Mechanistic models are often designed as a response to that subjectivity but, as we highlight here, they also produce another conflicting result or set of results; moreover, Bayesian model parameterization still introduces downstream variability, possibly even more so than niche modeling.

We also found that the mechanistic models we examined produced much more inclusive predictions than any other model we considered (in accordance with work in parallel fields similarly suggesting mechanistic models favor generality and realism over precision [315]). To some degree, this conflict may expose an underlying tension between two different intentions of disease mapping. One paradigm focuses on accuracy (especially specificity), and follows a similar paradigm to mainstream invasion biology research in that it attempts to most accurately project the final boundaries of incipient range expansion. Overprediction and underprediction are weighted as equal problems in this approach; the task of appropriate allocation of clinical resources is equally impeded by both margins of error. An alternative paradigm assumes that a Type II error (excluding regions at risk of outbreaks) is of far greater significance than a Type I error (predicting risk for areas that remain unaffected), from a preparedness standpoint; and reacts especially to the stakes of under-prediction by targeting predictions at any area that could, theoretically, sustain outbreaks. In reality, all disease distribution models fall somewhere on a continuum between the two paradigms, and modelers following best practices are likely to produce primarily objective results. But to the degree that no forecasting effort is fully unsupervised, and basic decisions (like including or excluding current outbreak data) introduce opportunities for subjectivity, conflict between these two approaches is likely to be an ongoing disciplinary problem beyond Zika.

We note also that deliberate choices we made in the epidemiological models we included similarly produced a specific, and extreme, result. By setting mosquito populations high and using wide stochastic priors rather than tailored Zika outbreak parameters, we simulated unrealistic outbreaks on a scale even greater than seen in Brazil. (However, we note that a landmark study just published estimates that 12.3 million cases of Zika are expected every year in Latin America and the Caribbean; and while the United States was unassessed, northern Mexico was identified as a region of high variability and therefore high epidemic potential. [316]) Our methods here are meant to illustrate the full potential of hidden uncertainty that epidemiological models might inherit from ecological assumptions. The range of projected epidemics varies among the ENM approaches by more than an order of magnitude, but even the smallest outbreak prediction is still five orders of magnitude higher than the case totals observed during the last outbreak season (223 real cases versus roughly 12 million simulated cases). In practice, epidemiological models fitted to data may absorb some of the uncertainty between different ecological forecasts if outbreaks are constrained in areas of disagreement by additional socioecological factors. The scale of the problem is difficult to evaluate except on a case by case basis; at a minimum, we conclude that understanding the epidemiological implications of ecological uncertainty is a key step towards improving ecologists' performance in pandemic preparedness.

Ecological niche modeling is a comparatively new statistical method in ecology, and it has only recently been applied to emerging infectious diseases. In under two decades, the statistical power of ENMs has grown exponentially, especially as increasingly complex methods for machine learning have been applied to the problem. The dozen or so methods currently employed offer a wide palette of options for potential modelers to choose from, and compounded with the wide range of potential environmental and social covariates, seemingly limitless combinations of possible models can be produced from a single dataset, each of which is statistically rigorous enough to be published. Although guidelines exist for method selection and model tuning (e.g., variable selection), tremendous user-end creativity is still possible. High-profile targets, such as vector-borne and other zoonotic diseases, frequently inspire conflicting models, but in mainstream species distribution modeling research, the impacts of those conflicts are often treated in as an academic problem. For infectious disease mapping, such conflict has conspicuous stakes that produce downstream uncertainty for stakeholders, clinicians, and policymakers.

Future Directions for the United States

Despite the disagreement between different modeling approaches and results, southern Florida and southern Texas clearly emerge across studies as the most at-risk regions of the continental United States for Zika virus outbreaks. This appears concordant with the broader consensus in public health research, especially given that these are already the only regions with a recent history of dengue outbreaks in the continental U.S. [317] We also note that, in many of the models we considered, Los Angeles county emerged as a potential area of significant concern, especially given its dense population. But for the rest of the country, model disagreement is high and unresolved.

Given the wide suitable area suggested by the majority of models, the low totals of autochthonous cases in the continental United States still seems surprising. Epidemiological work supports the idea that the 2016 outbreak was not anomalously small; recent work estimated the R_0 of the Miami-Dade outbreak in the low range of 0.5 to 0.8, and found that multiple introductions (an estimated 4 to 40) were a necessary precursor for an outbreak on the scale of the 256 cases in 2016. [317] Continued or larger outbreaks could be possible in the future if the high force of infection from traveler cases—which have so far been an order of magnitude more common in the U.S.—drives more significant outbreaks than the 2016 outbreak in Florida. More realistically, a number of factors likely prevent the United States from experiencing an outbreak on the scale that Brazil or Colombia experienced. Some are ecological; vector populations may be more strongly seasonal at higher latitudes, or the sylvatic cycle of Zika may be different in different parts of the Americas. The role non-human primates play in the transmission of Zika is still poorly understood [318, 319], but the absence of monkey hosts could plausibly limit transmission in the United States. Lessons from chikungunya suggest that attention may need to be paid to potential alternate, sylvatic vectors and associated hosts [320, 321], especially given the significant number of vectors that may be competent for Zika transmission in the United States. [300]

Other potential explanations for the limited spread of Zika through the United States are more social or socioecological in nature. In developed countries, household exposure is often secondary to outdoor exposure for *Aedes*, and in Miami-Dade county, it has been suggested that heterogeneity in outdoor exposure could have produced a much smaller, faster epidemic. [322] Other plausible explanations include better access to health care, preemptive vector control as part of Zika preparedness efforts, and significant fine-scale heterogeneity limiting mosquito populations in well populated areas (a factor that some models can accommodate [323], but niche models at the global scale do not). The last of these is most easily addressed through ecological tools, and finer-scale validation of downscaled ecological models is an important next step for ongoing forecasting. At the county scale, more detailed GIS data are needed to identify probable areas of suitable vector density; identifying those areas can reduce the population at risk (used to parameterize models) from the population of an entire county down to just those living in high-risk (or non-zero risk) areas.

Future Directions for Model Development

At the present time, the most common practice to address the ecology-epidemiology interface in the niche modeling literature is the use of population-at-risk (PAR) methods. Basic area-under-the-model population estimates are perhaps the simplest and most readily comparable possible epidemiological metric; only Messina *et al.* present a global PAR (2.17 billion people) based on their Zika virus niche model. Bogoch *et al.* revised that figure in a more regional assessment for Africa, Asia, and the Pacific that included traveler populations and a seasonal component to transmission, but to do so, substituted existing dengue models in place of actual Zika models. [305] If implemented more frequently, population-at-risk methods could be a simple *post hoc* way of comparing different ecological forecasts. However, these methods might accidentally introduce more alarm than they communicate risk (just as using susceptible populations, without any associated model of transmission, is a fairly uninformative proxy for an epidemic projection in mainstream epidemiology). The exercise carried out here illuminates one of the primary weaknesses of ecological niche modeling methods; namely, though ENMs have great value for defining the plausible outer bounds of transmission, they are largely unable to clarify the distribution of risk within these suitable areas (except in rare cases where extremely specific populations at risk can be measured, e.g., rural poor livestock keepers at risk of anthrax[324]). Modeling approaches that more directly interface ecological and epidemiological concepts of risk and hazard are perhaps the "Holy Grail" of work at this interface, and approaches along these lines have recently been tested for hemorrhagic viruses in Africa. [325, 326] But we show here that uncertainty and subjectivity on the ecological side are propagated through approaches like these, with no clear solution.

The uncertainty at this interface represents a major deficiency in our ability to forecast disease spread. However, there are a number of potential of avenues of exploration that may help improve efforts to directly link epidemiological forecasts and ecological projections. On the epidemiological side of the problem, travel-based models have shown promise for other

diseases [327], and have been applied in a limited capacity with dengue models to predict Zika risk. [305] These types of models can be applied with Zika-specific niche models for more detailed forecasts of traveler-driven outbreaks at the edges of suitability. But a more detailed epidemiological link is needed between traveler force of infection and the scale of subsequent local outbreaks; so far, that causation has only been investigated in reverse. [328] The role of sexual transmission also requires deeper investigation. Early work suggested sexual transmission might be a substantial factor explaining the explosive South American outbreak [329, 330], but recent work has suggested sexually-transmitted outbreaks are unlikely [331], even if sexual transmission increases the severity of vector-borne outbreaks [332]; others still argue these risks are "understated."[333, 334] Some work at the county level has already begun predicting Zika risk based on other sexually transmitted diseases [309], but for this to be useful to policymakers, a basic and accurate model of importance of sexual transmission is still needed. [335]

On the ecological side, consensus models (like the simple majority-rule model presented here) may be the first step towards decomposing suitability into something more epidemiologically-relevant. Development of alternative consensus models should aim to further clarify the level of suitability beyond the simple binomial categorization offered by ENM methods alone. The inclusion of a temporal component (i.e., the use of the conservative Mordecai projections of suitability for mosquitoes) enables some decomposition of the ENM results. The Mordecai *et al.* model illustrates that transmission is unlikely to be a year-round property of most areas, especially in temperate zones, and our exercise shows that reducing the months of possible transmission does significantly reduce total outbreak size. Time-specific ecological niche models have been used with great success to predict the dynamics of dengue [260], another *Aedes*-borne disease, and have been applied as a proxy for Zika risk. [305] However, these models will need to be developed specifically for Zika as more data become available, and time-specific ecological niche models will pose an additional challenge for consensus building with mechanistic time-sensitive models like Mordecai *et al.*'s. Finally, we suggest the frameworks underlying consensus models should be adaptable as additional occurrence data is made available. ENMs are typically presented as static instantiations of dynamic processes, whether they describe species ranges of the transmission niches of emerging infectious diseases. The ability of these models to contribute to our understanding of pathogens entering novel regions or hosts will hinge upon their flexibility in incorporating near-real-time data. [298] The computational frameworks for dynamic, updating niche models exist [336], but are an unexplored frontier in eco-epidemiology.

Data Availability Statement

Data generated or analyzed during this study are made available on Figshare: doi.org/10.6084/m9.figshare.5514961.v1

Acknowledgements

We thank Lewis Bartlett, Eva Harris, and others for helpful comments and guidance. This work was supported by the Rutgers University Center for Discrete Mathematics & Theoretical Computational Science (DIMACS) Mathematics of Planet Earth 2013+ Workshop on Zika & DIMACS MPE 2013+ Workshop on Appropriate Complexity Modeling of the Impacts of Global Change on Ecosystems, and the Centers for Disease Control and Prevention Epidemic Predictions Initiative (CDC EPI), a Center funded by NSF (EF-0553768). SJR was also supported by NSF DEB EEID 1518681, NSF DEB RAPID 1641145, and CDC grant 1U01CK000510-01: Southeastern Regional Center of Excellence in Vector-Borne Diseases: the Gateway Program. This publication was supported by the Cooperative Agreement Number above from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

Author contributions statement

CJC conceived of the study, and contributed the ecological niche models. SJR contributed the mechanistic models. ERD designed code for epidemiological simulations. CJC and ERD contributed to figures. All authors contributed to the writing and editing of the manuscript.

Competing interests statement

The authors declare no competing interests.

Table 5.1: A comparison of the different ecological forecasts. Four different methods, each performing well based on sufficient data and predictors, produce highly contrasting results. Out of a total of 3108 counties in the continental U.S., only five have experienced outbreaks (Cameron County, TX with 6 cases of local transmission in 2016; Miami-Dade, FL with 241; Palm Beach, FL with 8; Broward County, FL with 5; and Pinellas County, FL with 1)[337, 317]. Accuracy values were calculated from the confusion matrix of observed outbreaks against predicted suitability. The Carlson model comes closest to predicting the geography of those outbreaks most accurately; but all epidemiological models "overpredict" the number of suitable counties based on the current extent of outbreaks. (Mordecai results are split for the highest bound with minimum temperatures, and the lowest bound for maximum temperatures, to give the full range of predictions. Self reported AUC values are shown not as a comparative measure of accuracy, but simply as the self-reported accuracy of the studies. Samy *et al.* used the Partial ROC in place of the AUC but did not report values. NA = Not Applicable; NR = Not Reported)

	Carlson	Messina	Samy
n_{points}	242	323	168
$n_{predictors}$	15	6	15
AUC	0.970	0.829	NA
Counties Predicted	13	465	1616
Accuracy	99.6%	85.2%	48.2%
County Population at Risk	19,653,445	95,359,408	270,249,781
Mean Outbreak Size	12,871,005	63,622,367	181,290,371
Median Outbreak Size	14,552,250	64,038,273	181,732,629

	Mordecai (97.5% min)	Mordecai (2.5% max)
n_{points}	NA	NA
$n_{predictors}$	NA	NA
AUC	NA	NA
Counties Predicted	1937	3108
Accuracy	37.8%	0.2%
County Population at Risk	218,444,263	320,957,062
Mean Outbreak Size	37,598,099	198,910,979
Median Outbreak Size	37,312,233	197,731,918

Table 5.2: Aggregating risk to the county scale can absorb some of the inherent spatial uncertainty of ecological niche modeling, but is itself an assumption that changes downstream impacts on the scale of outbreaks, as well as the scale of disagreement between models.

	Carlson	Partial	Messina	Partial
Counties Predicted	13	13	465	465
County Population at Risk	19,653,445	6,264,516	95,359,408	38,085,602
Mean Outbreak Size	12,871,005	4,195,326	63,622,367	25,897,671
Median Outbreak Size	14,552,250	4,262,636	64,038,273	26,307,445

	Samy	Partial
Counties Predicted	1616	1616
County Population at Risk	270,249,781	99,324,226
Mean Outbreak Size	181,290,371	66,345,567
Median Outbreak Size	181,732,629	66,850,610

Table 5.3: Outbreak simulations exhibit greater than threefold variation in predictions among the four models presented in Samy.

	Model 1	Model 2	Model 3	Model 4
Counties Predicted	338	2197	670	1616
County Population at Risk	91,174,791	296,007,551	148,700,587	270,249,781
Mean Outbreak Size	59,561,603	197,553,462	98,314,605	181,290,371
Median Outbreak Size	59,561,222	197,759,983	98,770,516	181,732,629

Table 5.4: Majority rule based consensus models, meant to resolve uncertainty between the forecasts and provide a middle scenario. The main majority rule model combines the Carlson, Messina, and Samy forecasts; the seasonal majority rule model assigns monthly suitability values to that forecast, based on the minimum temperature 97.5% Mordecai model.

	Majority Rule	Seasonal Majority Rule
Counties Predicted	383	370
County Population at Risk	93,195,970	87,632,865
Mean Outbreak Size	60,259,904	24,267,441
Median Outbreak Size	60,940,111	24,407,885

Table 5.5: Epidemiological parameters for the SEIIAR & SEI models presented in the main text; values taken directly from from Table 1 of Gao *et al.*'s study [312].

Parameter	Description	Range
a	Mosquito biting rate ($mosquito^{-1}day^{-1}$)	$(0.3, 1)$
b	Mosquito to human transmission rate ($bite^{-1}$)	$(0.1, 0.75)$
c	Human to mosquito transmission rate ($bite^{-1}$)	$(0.3, 0.75)$
β	Human to human (sexual) transmission rate	$(0.001, 0.1)$
$1/\gamma_{h1}$	Infectious period (acute)	$(3, 7)$
$1/\gamma_{h2}$	Infectious period (convalescent)	$(14, 30)$
$1/\gamma_h$	Infectious period (asymptomatic)	$(5, 10)$
η	Exposed human to mosquito transmission proportion	$(5, 10)$
θ	Proportion symptomatic infections	$(0.1, 0.27)$
κ	Exposed human to human transmission proportion	$(0, 1)$
$1/\mu$	Mosquito lifespan (day)	$(4, 35)$
$1/\nu_h$	Intrinsic incubation period (day)	$(2, 7)$
$1/\nu_v$	Extrinsic incubation period (day)	$(8, 12)$
τ	Convalescent human to human transmission proportion	$(0, 1)$

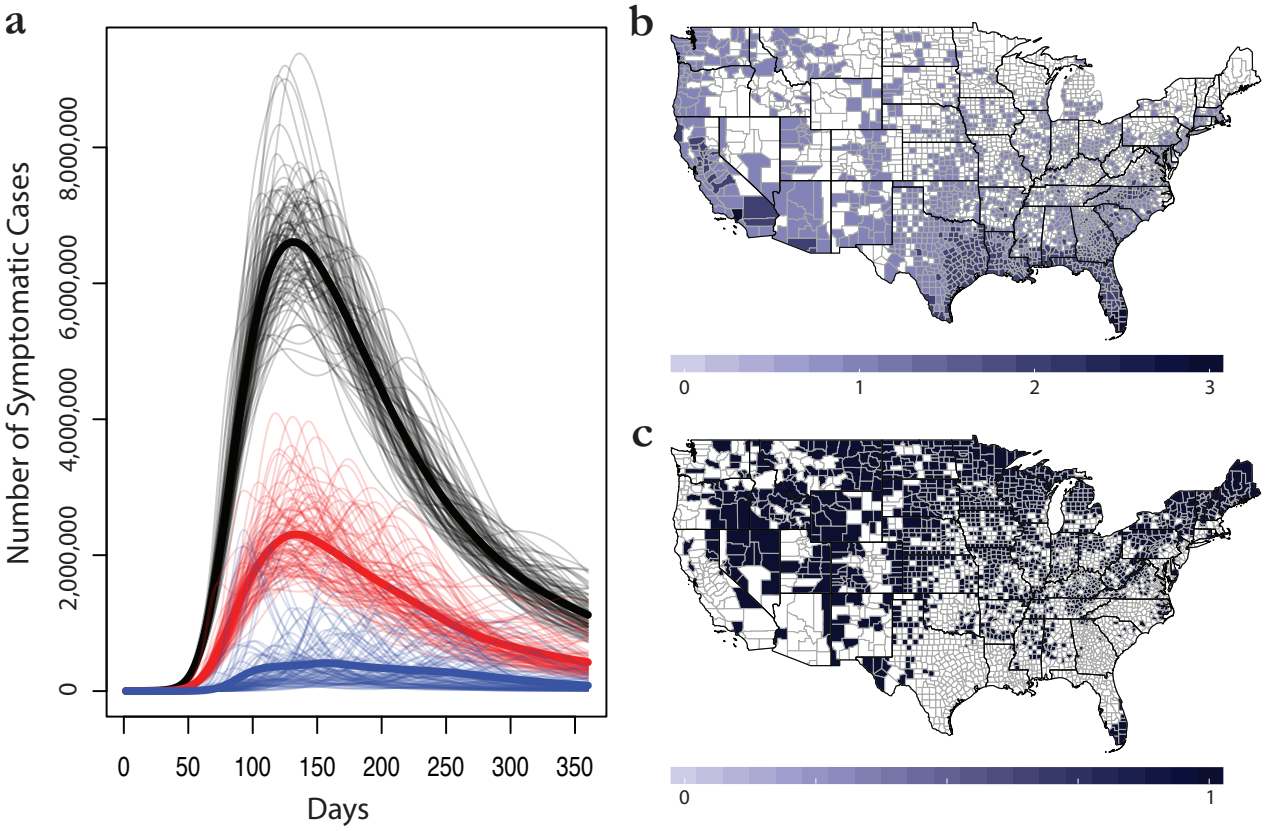


Figure 5.1: The margin of error in ecological niche models for Zika virus. (a) Average epidemiological forecasts associated with county data for Carlson (*blue*), Messina (*red*), and Samy (*black*), against a backdrop of overlapping individual simulations for each (grey). (b) The individual predictions of each model are given as presence or absence values; a maximum score of 3 indicates all models agree on presence, while a score of 0 indicates all models agree on absence. (c) Has consensus been achieved? At the county scale, dark blue indicates consensus among niche models; white indicates controversy. Maps were made in R 3.3.2 [338], using U.S. Census shapefiles.

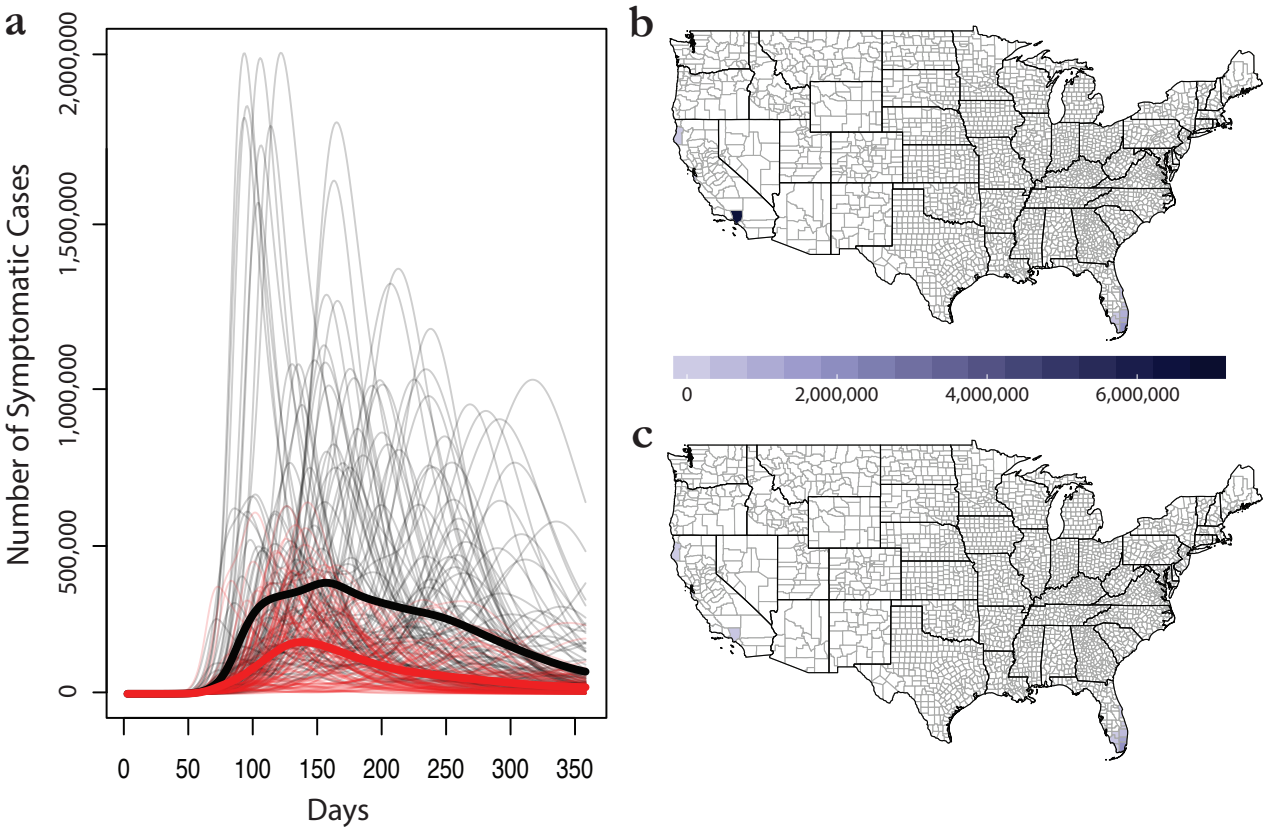


Figure 5.2: Full county versus partial population simulations with the Carlson model. Outbreak simulations (a) are given in black for the full county, and red for the partial county. Mean county case totals are mapped for the full county (b) and partial county (c).

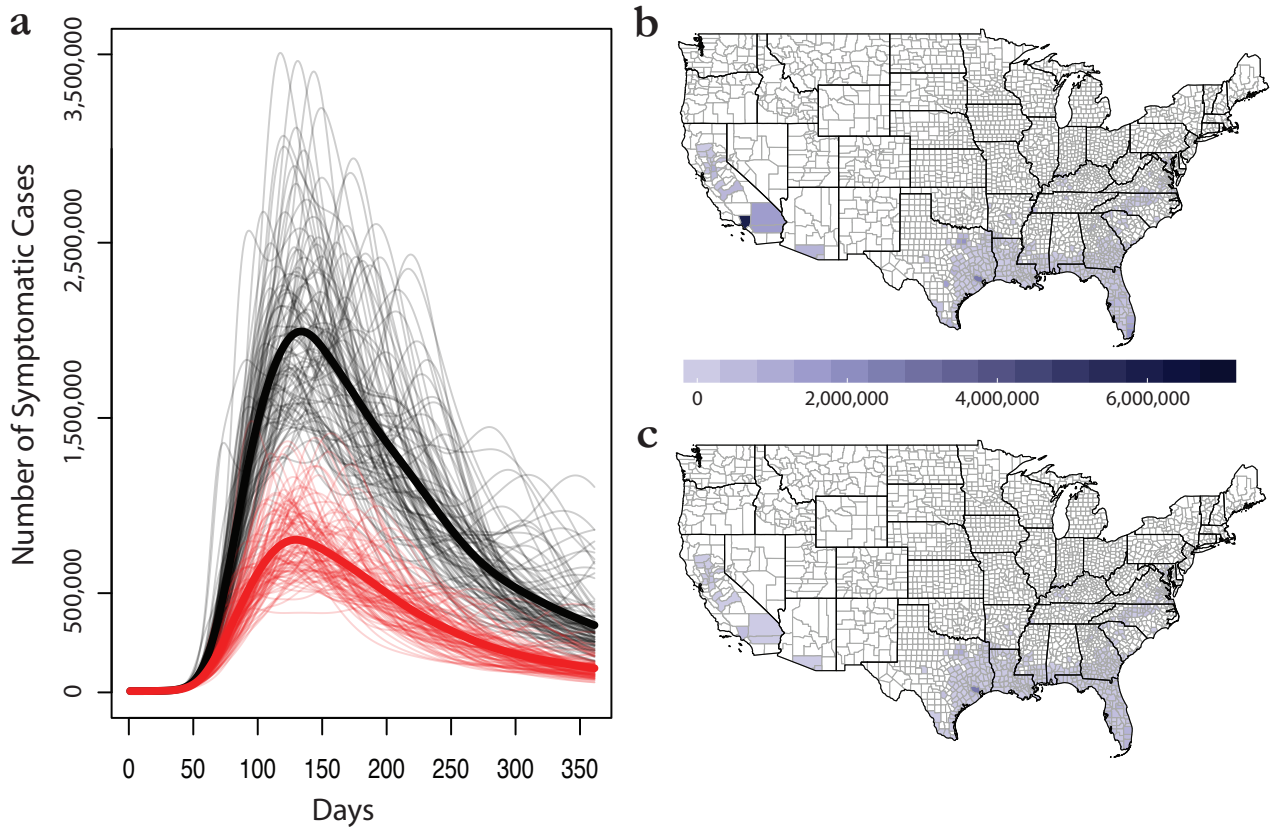


Figure 5.3: Full county versus partial population simulations with the Messina model. Outbreak simulations (a) are given in black for the full county, and red for the partial county. Mean county case totals are mapped for the full county (b) and partial county (c).

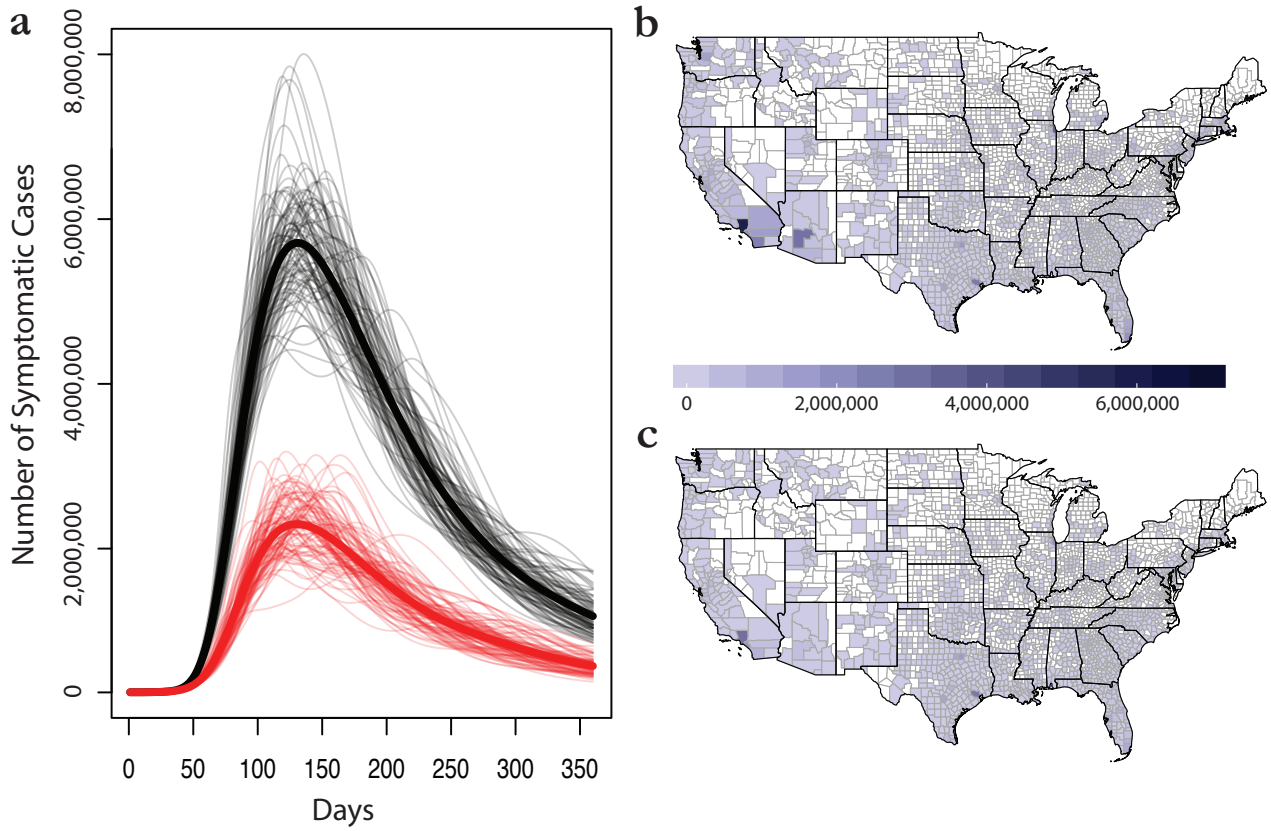


Figure 5.4: Full county versus partial population simulations with the Samy model. Outbreak simulations (a) are given in black for the full county, and red for the partial county. Mean county case totals are mapped for the full county (b) and partial county (c).

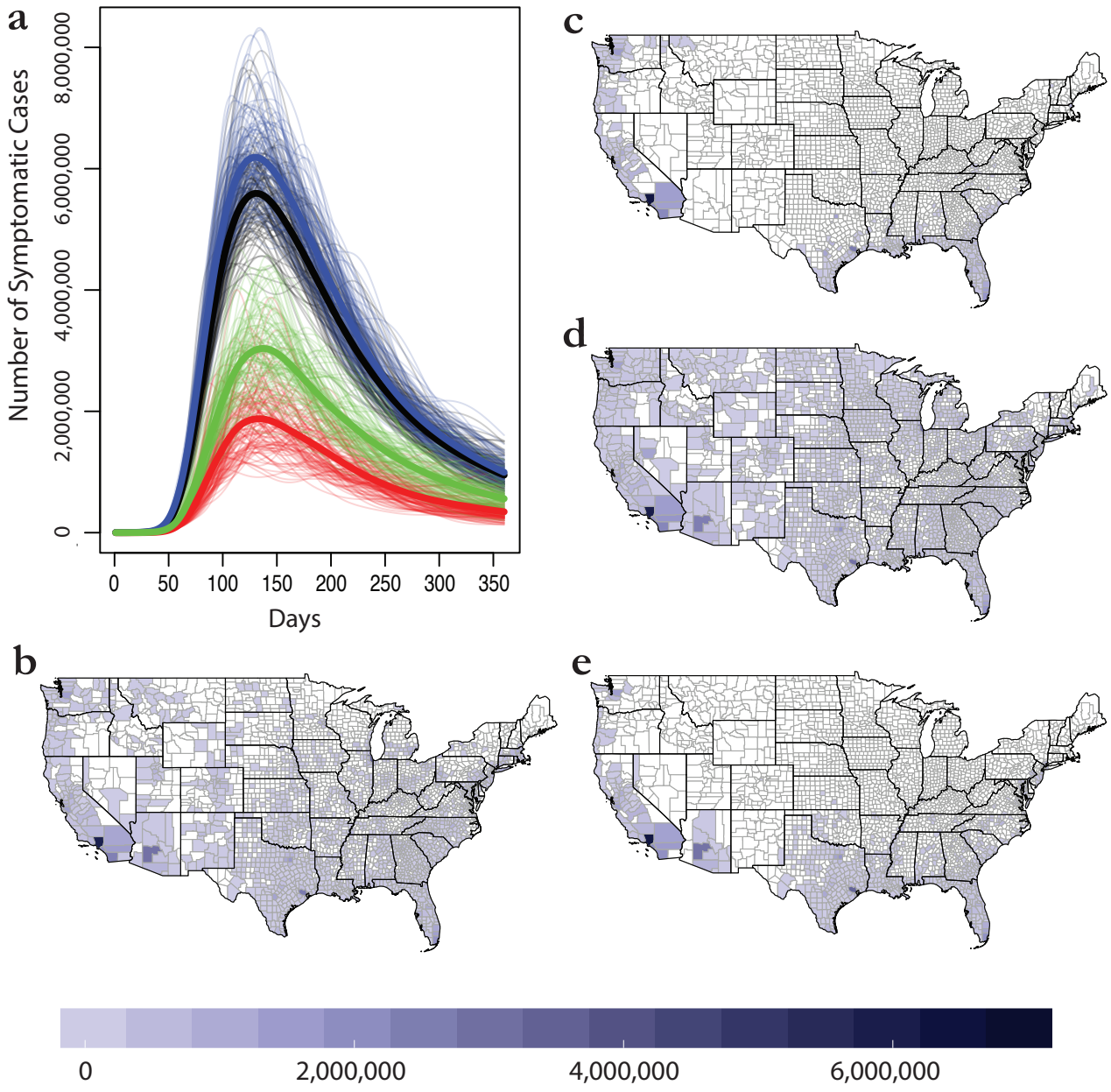


Figure 5.5: Variation within the Samy models. Outbreak trajectories are shown in (a) for models 1 (red), 2 (blue), 3 (green), and 4 (black). Bolded lines are mean trajectories. Final average case totals are then mapped for model 4 (b), the main model we discuss in the text and use in other comparisons, as well as models 1 (c), 2 (d), and 3 (e).

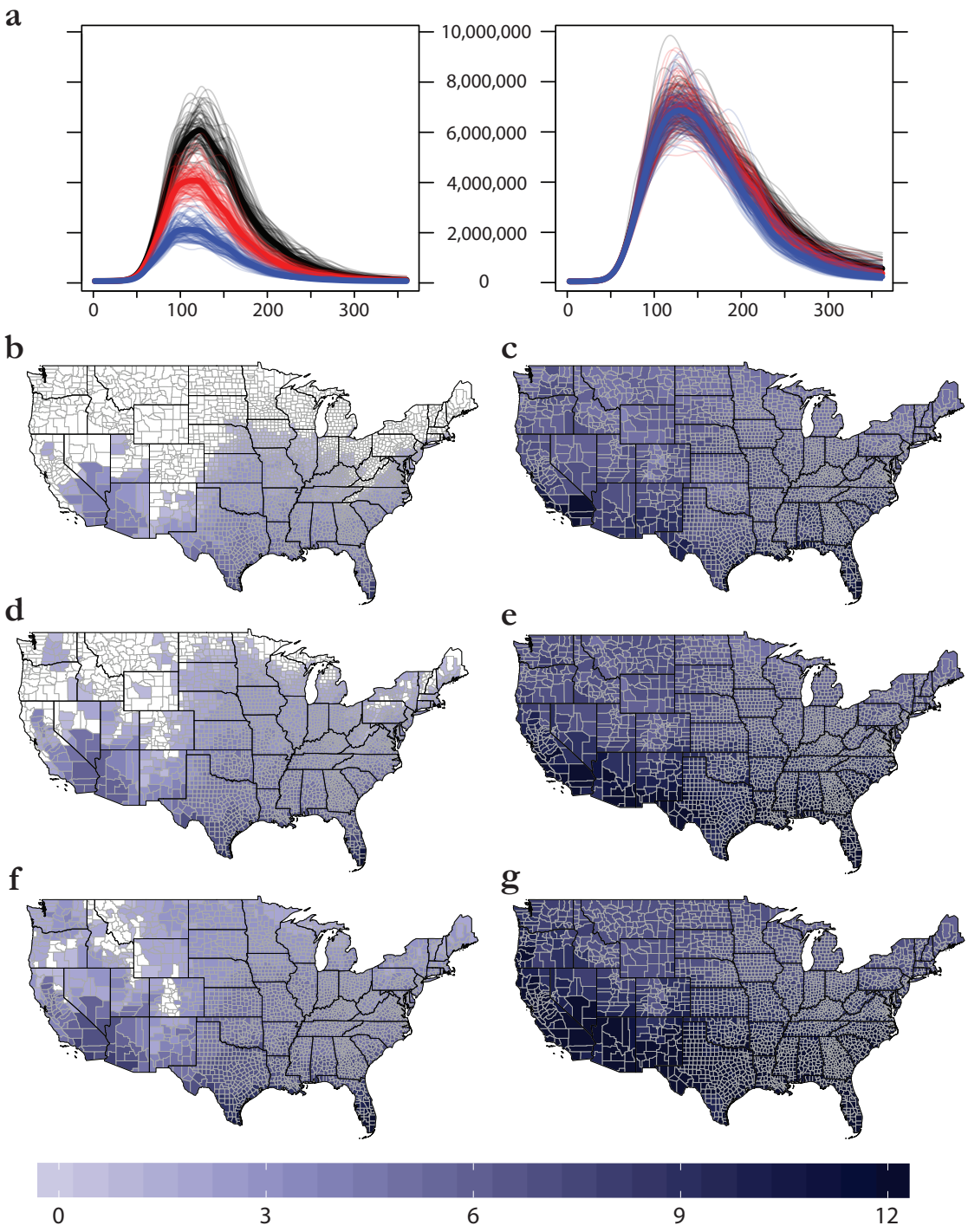


Figure 5.6: The margin of error within a single Bayesian mechanistic model for Zika virus, applied to minimum (left) and maximum (right) monthly temperatures. (a) 100 outbreak simulations for 97.5% (blue), 50% (red), and 2.5% (black) confidence intervals. (b-f) The number of months each county is predicted to be suitable for Zika virus transmission ($R_0 > 0$) for 97.5% (b,c), 50% (d,e), and 2.5% (f,g) scenarios. Maps were made in R 3.3.2 [338], using U.S. Census shapefiles.

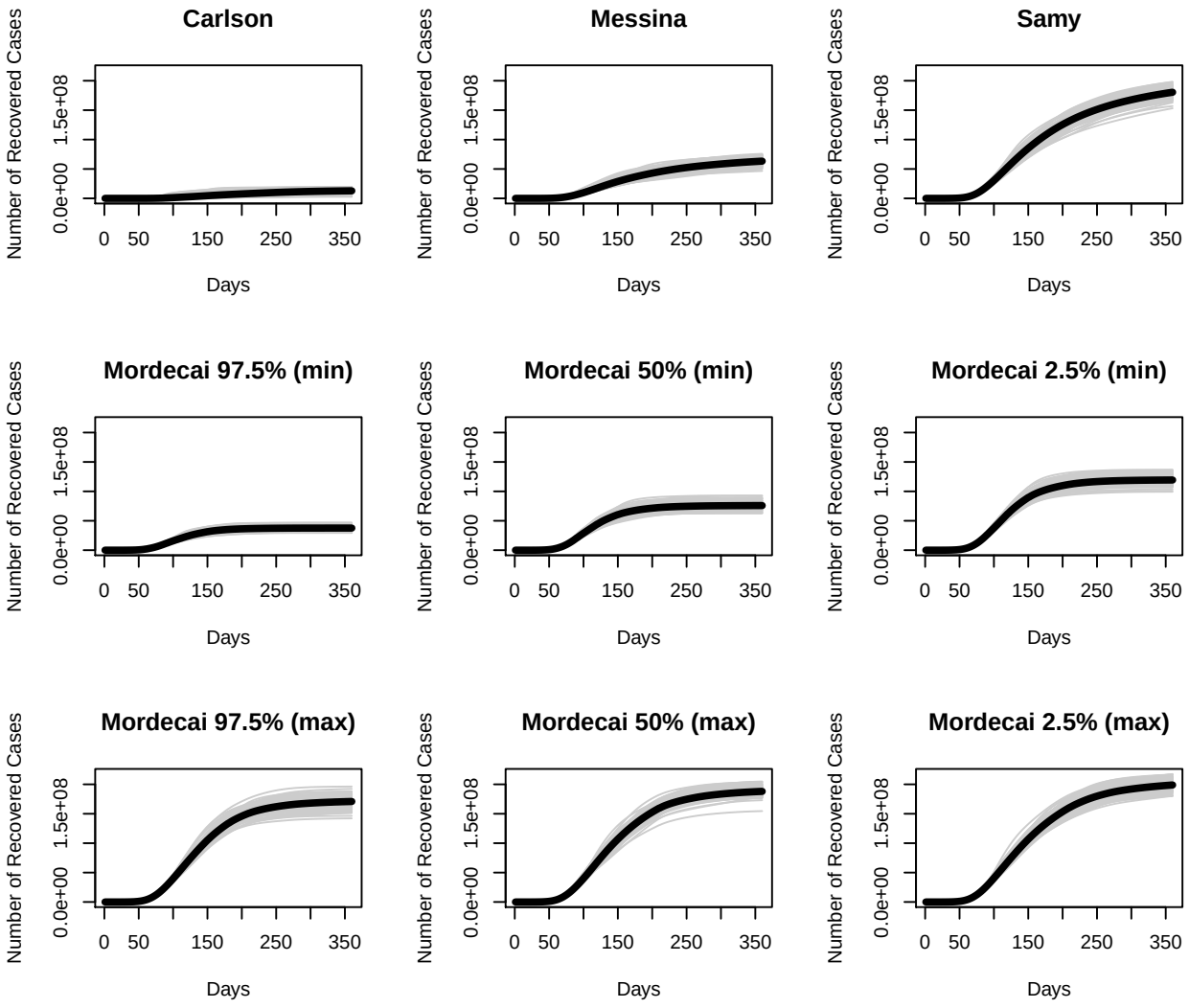


Figure 5.7: Nine possible trajectories for outbreaks in the United States: three based on ecological niche models, and six based on Bayesian mechanistic forecasts. (*y-axis on log scale*)

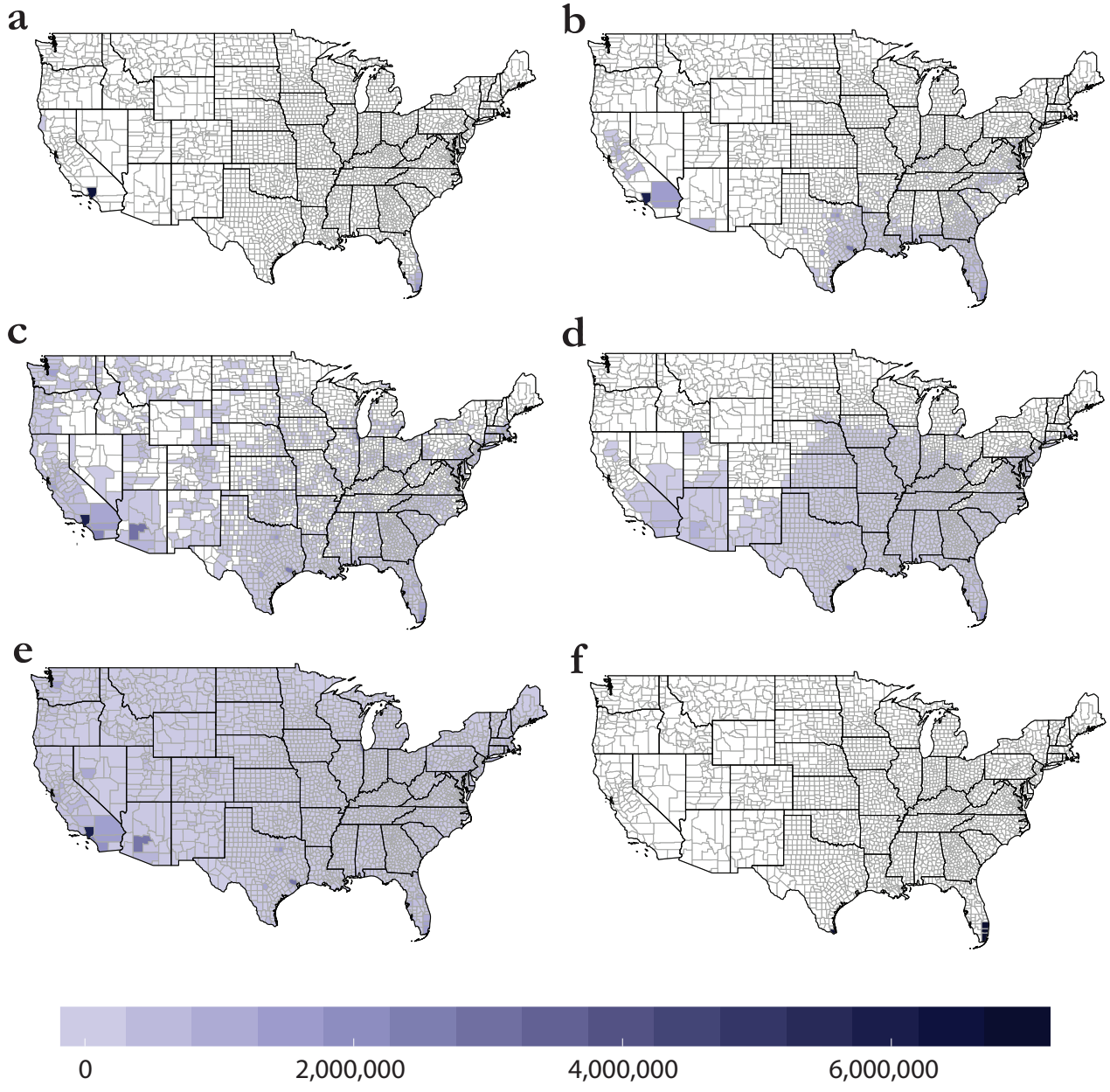


Figure 5.8: Case totals by county for (a) Carlson, (b) Messina, (c) Samy, (d), Mordecai 97.5% confidence (minimum temperatures), and (e) Mordecai 2.5% confidence (max temperatures), compared against (f) counties with reported autochthonous transmission in 2016 (three in Florida, one in Texas). Maps were made in R 3.3.2 [338], using U.S. Census shapefiles.

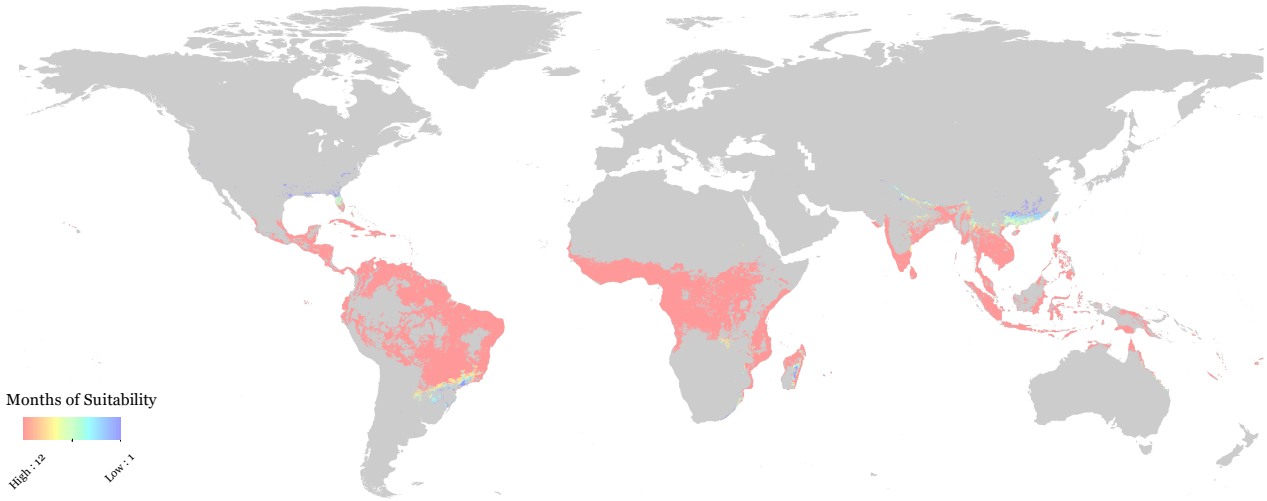


Figure 5.9: A global, consensus-based, seasonal (monthly) majority rule map of suitability for Zika virus transmission.

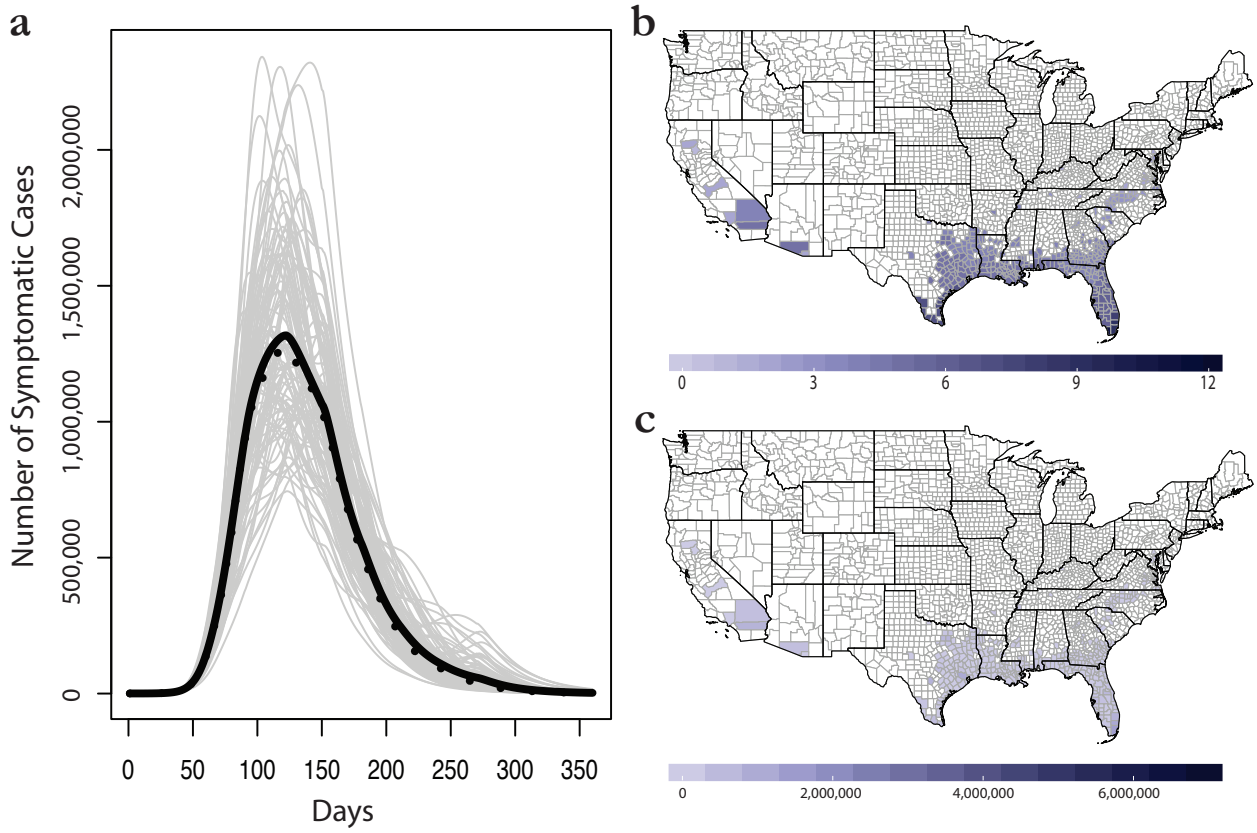


Figure 5.10: The seasonal majority rule method for consensus building across ecological forecasts. (a) Mean (black) and median (dashed) trajectories for 100 epidemic simulations. (b) The majority rule map: shading represents the number of months each county is marked suitable for outbreaks. (c) Final average case totals in the seasonal majority rule method. Maps were made in R 3.3.2 [338], using U.S. Census shapefiles.

Bibliography

[1] Carrie A Cizauskas et al. "Parasite vulnerability to climate change: an evidence-based functional trait approach". In: *Royal Society Open Science* 4.1 (2017), p. 160535.

[2] Tad Dallas. "helminthR: an R interface to the London Natural History Museum's host–parasite database". In: *Ecography* 39.4 (2016), pp. 391–393.

[3] Andy Dobson et al. "Homage to Linnaeus: How many parasites? How many hosts?" In: *Proceedings of the National Academy of Sciences* 105.Supplement 1 (2008), pp. 11482–11489.

[4] Brendan B Larsen et al. "Inordinate Fondness Multiplied and Redistributed: the Number of Species on Earth and the New Pie of Life". In: *The Quarterly Review of Biology* 92.3 (2017), pp. 229–265.

[5] Eric R Dougherty et al. "Paradigms for parasite conservation". In: *Conservation Biology* 30.4 (2016), pp. 724–733.

[6] Takuya Sato et al. "Nematomorph parasites drive energy flow through a riparian ecosystem". In: *Ecology* 92.1 (2011), pp. 201–207.

[7] Takuya Sato et al. "Nematomorph parasites indirectly alter the food web and ecosystem function of streams through behavioural manipulation of their cricket hosts". In: *Ecology Letters* 15.8 (2012), pp. 786–793.

[8] Chris D Thomas et al. "Extinction risk from climate change". In: *Nature* 427.6970 (2004), pp. 145–148.

[9] A Townsend Peterson. *Mapping disease transmission risk: enriching models using biogeography and ecology*. JHU Press, 2014.

[10] Pieter TJ Johnson et al. "Host and parasite diversity jointly control disease risk in complex communities". In: *Proceedings of the National Academy of Sciences* 110.42 (2013), pp. 16916–16921.

[11] Abdallah M Samy et al. "Mapping the global geographic potential of Zika virus spread". In: *Memórias do Instituto Oswaldo Cruz* 111.9 (2016), pp. 559–560.

[12] Jane P Messina et al. "Mapping global environmental suitability for Zika virus". In: *Elife* 5 (2016), e15272.

[13] Patrick R Stephens et al. "The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distributions and impacts". In: *Ecology letters* 19.9 (2016), pp. 1159–1171.

[14] Pieter TJ Johnson, Jacobus C De Roode, and Andy Fenton. "Why infectious disease research needs community ecology". In: *Science* 349.6252 (2015), p. 1259504.

[15] Camilo Mora et al. "How many species are there on Earth and in the ocean?" In: *PLoS Biol* 9.8 (2011), e1001127.

[16] David M Raup and J John Sepkoski Jr. "Mass extinctions in the marine fossil record". In: *Science* 215.4539 (1982), pp. 1501–1503.

[17] Michael L McKinney and Julie L Lockwood. "Biotic homogenization: a few winners replacing many losers in the next mass extinction". In: *Trends in ecology & evolution* 14.11 (1999), pp. 450–453.

[18] Elizabeth Kolbert. *The sixth extinction: An unnatural history*. A&C Black, 2014.

[19] Gerardo Ceballos et al. "Accelerated modern human–induced species losses: Entering the sixth mass extinction". In: *Science advances* 1.5 (2015), e1400253.

[20] Anthony D Barnosky et al. "Has the Earth's sixth mass extinction already arrived?" In: *Nature* 471.7336 (2011), pp. 51–57.

[21] R Levins. "Extinction". In: *Lectures on Mathematics in the Life Sciences* 2 (1970), pp. 77–107.

[22] Norman Myers. *The sinking ark*. Pergamon Press, Oxford, 1979.

[23] John M Drake. "Extinction times in experimental populations". In: *Ecology* 87.9 (2006), pp. 2215–2220.

[24] John M Drake. "Tail probabilities of extinction time in a large number of experimental populations". In: *Ecology* 95.5 (2014), pp. 1119–1126.

[25] Graeme Caughley. "Directions in conservation biology". In: *Journal of animal ecology* (1994), pp. 215–244.

[26] Mark S Boyce. "Population growth with stochastic fluctuations in the life table". In: *Theoretical Population Biology* 12.3 (1977), pp. 366–373.

[27] Steinar Engen and Bernt-Erik Sæther. "Predicting the time to quasi-extinction for populations far below their carrying capacity". In: *Journal of theoretical Biology* 205.4 (2000), pp. 649–658.

[28] Bernt-Erik Sæther and Steinar Engen. "Including uncertainties in population viability analysis using population prediction intervals". In: *Population viability analysis* (2002), pp. 191–212.

[29] Steven R Beissinger. "Population viability analysis: past, present, future". In: *Population viability analysis* (2002), pp. 5–17.

[30] Russell Lande. "Risks of population extinction from demographic and environmental stochasticity and random catastrophes". In: *American Naturalist* (1993), pp. 911–927.

[31] Michael E Gilpin. "Minimum viable populations: processes of species extinction". In: *Conservation biology: the science of scarcity and diversity* (1986), pp. 19–34.

[32] Barry W Brook, Lochran W Traill, and Corey JA Bradshaw. "Minimum viable population sizes and global extinction risk are unrelated". In: *Ecology letters* 9.4 (2006), pp. 375–382.

[33] Sewall Wright. "Isolation by distance under diverse systems of mating". In: *Genetics* 31.1 (1946), p. 39.

[34] Sewall Wright. "The interpretation of population structure by F-statistics with special regard to systems of mating". In: *Evolution* (1965), pp. 395–420.

[35] Brian Charlesworth and Deborah Charlesworth. "The Evolutionary Effects of Finite Population Size: Basic Theory". In: *Elements of Evolutionary Genetics*. 2012. Chap. 5, pp. 195–244.

[36] Fred W Allendorf and Nils Ryman. "The role of genetics in population viability analysis". In: *Population viability analysis. University of Chicago Press, Chicago* (2002), pp. 50–85.

[37] Warren E Johnson et al. "Genetic restoration of the Florida panther". In: *Science* 329.5999 (2010), pp. 1641–1645.

[38] Hal Caswell. *Matrix population models*. Wiley Online Library, 2001.

[39] Wayne M Getz and Robert G Haight. *Population harvesting: demographic models of fish, forest, and animal resources*. Vol. 27. Princeton University Press, 1989.

[40] Wayne M Getz. "A hypothesis regarding the abruptness of density dependence and the growth rate of populations". In: *Ecology* 77.7 (1996), pp. 2014–2026.

[41] Brett A Melbourne and Alan Hastings. "Extinction risk depends strongly on factors contributing to stochasticity". In: *Nature* 454.7200 (2008), pp. 100–103.

[42] Wayne M Getz et al. "A web app for population viability and harvesting analyses". In: *Natural Resource Modeling* (2016).

[43] Stephen P Ellner et al. "Precision of population viability analysis". In: *Conservation Biology* 16.1 (2002), pp. 258–261.

[44] J Michael Reed et al. "Emerging issues in population viability analysis". In: *Conservation biology* 16.1 (2002), pp. 7–19.

[45] Tim Coulson et al. "The use and abuse of population viability analysis". In: *Trends in Ecology & Evolution* 16.5 (2001), pp. 219–221.

[46] Steven R Beissinger and M Ian Westphal. "On the use of demographic models of population viability in endangered species management". In: *The Journal of wildlife management* (1998), pp. 821–841.

[47] EJ Milner-Gulland. "Catastrophe and hope for the saiga". In: *Oryx* 49.04 (2015), pp. 577–577.

[48] Francisco De Castro and Benjamin Bolker. "Mechanisms of disease-induced extinction". In: *Ecology Letters* 8.1 (2005), pp. 117–126.

[49] SA Frank and P Schmid-Hempel. "Mechanisms of pathogenesis and the evolution of parasite virulence". In: *Journal of evolutionary biology* 21.2 (2008), pp. 396–404.

[50] Matthew C Fisher et al. "Emerging fungal threats to animal, plant and ecosystem health". In: *Nature* 484.7393 (2012), pp. 186–194.

[51] Deanna H Olson et al. "Mapping the global emergence of Batrachochytrium dendrobatidis, the amphibian chytrid fungus". In: *PloS one* 8.2 (2013), e56802.

[52] Lee Francis Skerratt et al. "Spread of chytridiomycosis has caused the rapid global decline and extinction of frogs". In: *EcoHealth* 4.2 (2007), pp. 125–134.

[53] J Alan Pounds et al. "Widespread amphibian extinctions from epidemic disease driven by global warming". In: *Nature* 439.7073 (2006), pp. 161–167.

[54] Malcolm L McCallum. "Amphibian decline or extinction? Current declines dwarf background extinction rate". In: *Journal of Herpetology* 41.3 (2007), pp. 483–491.

[55] Winifred F Frick et al. "An emerging disease causes regional population collapse of a common North American bat species". In: *Science* 329.5992 (2010), pp. 679–682.

[56] Sonia Altizer, Drew Harvell, and Elizabeth Friedle. "Rapid evolutionary dynamics and disease threats to biodiversity". In: *Trends in Ecology & Evolution* 18.11 (2003), pp. 589–596.

[57] Brooke Maslo and Nina H Fefferman. "A case study of bats and white-nose syndrome demonstrating how to model population viability with evolutionary effects". In: *Conservation Biology* 29.4 (2015), pp. 1176–1185.

[58] Karen C Abbott. "Does the pattern of population synchrony through space reveal if the Moran effect is acting?" In: *Oikos* 116.6 (2007), pp. 903–912.

[59] Karen C Abbott. "A dispersal-induced paradox: synchrony and stability in stochastic metapopulations". In: *Ecology letters* 14.11 (2011), pp. 1158–1169.

[60] Ilkka Hanski. "Metapopulation dynamics". In: *Nature* 396.6706 (1998), pp. 41–49.

[61] James C Bull et al. "Metapopulation extinction risk is increased by environmental stochasticity and assemblage complexity". In: *Proceedings of the Royal Society of London B: Biological Sciences* 274.1606 (2007), pp. 87–96.

[62] Ana R Gouveia, Ottar N Bjørnstad, and Emil Tkadlec. "Dissecting geographic variation in population synchrony using the common vole in central Europe as a test bed". In: *Ecology and evolution* 6.1 (2016), pp. 212–218.

[63] Richard Levins. "Some demographic and genetic consequences of environmental heterogeneity for biological control". In: *Bulletin of the Entomological society of America* 15.3 (1969), pp. 237–240.

[64] F Elías-Wolff et al. "How Levins' dynamics emerges from a Ricker metapopulation model". In: *Theoretical Ecology* 2.9 (2016), pp. 173–183.

[65] Ilkka Hanski. "Single-species metapopulation dynamics: concepts, models and observations". In: *Biological Journal of the Linnean Society* 42.1-2 (1991), pp. 17–38.

[66] Luis J Gilarranz and Jordi Bascompte. "Spatial network structure and metapopulation persistence". In: *Journal of Theoretical Biology* 297 (2012), pp. 11–16.

[67] Robert H MacArthur and Edward O Wilson. *Theory of Island Biogeography.(MPB-1)*. Vol. 1. Princeton University Press, 2015.

[68] Ilkka Hanski and Otso Ovaskainen. "The metapopulation capacity of a fragmented landscape". In: *Nature* 404.6779 (2000), pp. 755–758.

[69] Jacopo Grilli, György Barabás, and Stefano Allesina. "Metapopulation persistence in random fragmented landscapes". In: *PLoS Comput Biol* 11.5 (2015), e1004251.

[70] Subhashni Taylor et al. "Applications of Rapid Evaluation of Metapopulation Persistence (REMP) in Conservation Planning for Vulnerable Fauna Species". In: *Environmental management* 57.6 (2016), pp. 1281–1291.

[71] Manojit Roy, Robert D Holt, and Michael Barfield. "Temporal autocorrelation can enhance the persistence and abundance of metapopulations comprised of coupled sinks". In: *The American Naturalist* 166.2 (2005), pp. 246–261.

[72] V AA Jansen and Jin Yoshimura. "Populations can persist in an environment consisting of sink habitats only". In: *Proc. Natl. Acad. Sci. USA* 95 (1998), pp. 3696–3698.

[73] David P Matthews and Andrew Gonzalez. "The inflationary effects of environmental fluctuations ensure the persistence of sink metapopulations". In: *Ecology* 88.11 (2007), pp. 2848–2856.

[74] Michael B Bonsall and Alan Hastings. "Demographic and environmental stochasticity in predator–prey metapopulation dynamics". In: *Journal of Animal Ecology* 73.6 (2004), pp. 1043–1055.

[75] Mathew A Leibold et al. "The metacommunity concept: a framework for multi-scale community ecology". In: *Ecology letters* 7.7 (2004), pp. 601–613.

[76] Richard Gomulkiewicz and Robert D Holt. "When does evolution by natural selection prevent extinction?" In: *Evolution* 49.1 (1995), pp. 201–207.

[77] Russell Lande. "Natural selection and random genetic drift in phenotypic evolution". In: *Evolution* (1976), pp. 314–334.

[78] Luis-Miguel Chevin, Russell Lande, and Georgina M Mace. "Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory". In: *PLoS Biol* 8.4 (2010), e1000357.

[79] Carl D Schlichting, Massimo Pigliucci, et al. *Phenotypic evolution: a reaction norm perspective.* Sinauer Associates Incorporated, 1998.

[80] Yi-Qi Hao et al. "Evolutionary rescue can be impeded by temporary environmental amelioration". In: *Ecology letters* 18.9 (2015), pp. 892–898.

[81] Graham Bell and Andrew Gonzalez. "Adaptation and evolutionary rescue in metapopulations experiencing environmental deterioration". In: *Science* 332.6035 (2011), pp. 1327–1330.

[82] Haley A Lindsey et al. "Evolutionary rescue from extinction is contingent on a lower rate of environmental change". In: *Nature* 494.7438 (2013), pp. 463–467.

[83] Katja Schiffers et al. "Limited evolutionary rescue of locally adapted populations facing climate change". In: *Philosophical transactions of the Royal Society of London B: Biological sciences* 368.1610 (2013), p. 20120083.

[84] Barry W Brook, Navjot S Sodhi, and Corey JA Bradshaw. "Synergies among extinction drivers under global change". In: *Trends in ecology & evolution* 23.8 (2008), pp. 453–460.

[85] Lewis J Bartlett et al. "Synergistic impacts of habitat loss and fragmentation on model ecosystems". In: *Proc. R. Soc. B.* Vol. 283. 1839. The Royal Society. 2016, p. 20161027.

[86] Catalina Pimiento and Christopher F Clements. "When did Carcharocles megalodon become extinct? A new analysis of the fossil record". In: *PLoS One* 9.10 (2014), e111086.

[87] David L Roberts and Andrew R Solow. "Flightless birds: when did the dodo become extinct?" In: *Nature* 426.6964 (2003), pp. 245–245.

[88] NJ Collar. "Extinction by assumption; or, the Romeo Error on Cebu". In: *Oryx* 32.4 (1998), pp. 239–244.

[89] Brett R Scheffers et al. "The world's rediscovered species: back from the brink?" In: *PloS one* 6.7 (2011), e22531.

[90] DS Robson and JH Whitlock. "Estimation of a truncation point". In: *Biometrika* 51.1/2 (1964), pp. 33–39.

[91] Andrew R Solow. "Inferring extinction from sighting data". In: *Ecology* 74.3 (1993), pp. 962–964.

[92] Andrew R Solow. "Inferring extinction from a sighting record". In: *Mathematical biosciences* 195.1 (2005), pp. 47–55.

[93] Christopher F Clements et al. "Experimentally testing the accuracy of an extinction estimator: Solow's optimal linear estimation model". In: *Journal of Animal Ecology* 82.2 (2013), pp. 345–354.

[94] Andrew R Solow and Andrew R Beet. "On uncertain sightings and inference about extinction". In: *Conservation Biology* 28.4 (2014), pp. 1119–1123.

[95] Chris S Elphick, David L Roberts, and J Michael Reed. "Estimated dates of recent extinctions for North American and Hawaiian birds". In: *Biological Conservation* 143.3 (2010), pp. 617–624.

[96] Tony Juniper. *Spix's Macaw: the race to save the world's rarest bird*. Simon and Schuster, 2004.

[97] C Clements. *sExtinct: Calculates the historic date of extinction given a series of sighting events. R package version 1.1*. 2013.

[98] John W Fitzpatrick et al. "Ivory-billed Woodpecker (Campephilus principalis) persists in continental North America". In: *Science* 308.5727 (2005), pp. 1460–1462.

[99] David L Roberts, Chris S Elphick, and J Michael Reed. "Identifying anomalous reports of putatively extinct species and why it matters". In: *Conservation Biology* 24.1 (2010), pp. 189–196.

[100] Tamsin E Lee et al. "Inferring extinctions from sighting records of variable reliability". In: *Journal of applied ecology* 51.1 (2014), pp. 251–258.

[101] CJ Thompson et al. "Inferring extinction risks from sighting records". In: *Journal of theoretical biology* 338 (2013), pp. 16–22.

[102] Andrew R Solow. "On the prior distribution of extinction time". In: *Biology Letters* 12.6 (2016), p. 20160089.

[103] Elizabeth H Boakes, Tracy M Rout, and Ben Collen. "Inferring species extinction: the use of sighting records". In: *Methods in Ecology and Evolution* 6.6 (2015), pp. 678–687.

[104] Tamsin E Lee, Clive Bowman, and David L Roberts. "Are extinction opinions extinct?" In: *PeerJ* 5 (2017), e3663.

[105] Nicholas J Gotelli et al. "Specimen-Based Modeling, Stopping Rules, and the Extinction of the Ivory-Billed Woodpecker". In: *Conservation Biology* 26.1 (2012), pp. 47–56.

[106] David A Sibley et al. "Ivory-billed or pileated woodpecker?" In: *Science* 315.5818 (2007), pp. 1495–1496.

[107] Jonathon C Dunn et al. "Mapping the potential distribution of the Critically Endangered Himalayan Quail Ophrysia superciliosa using proxy species and species distribution modelling". In: *Bird Conservation International* 25.4 (2015), pp. 466–478.

[108] JL McCune. "Species distribution models predict rare species occurrences despite significant effects of landscape context". In: *Journal of Applied Ecology* 53.6 (2016), pp. 1871–1879.

[109] Huijie Qiao, Jorge Soberón, and Andrew Townsend Peterson. "No silver bullets in correlative ecological niche modelling: insights from testing among many potential algorithms for niche estimation". In: *Methods in Ecology and Evolution* 6.10 (2015), pp. 1126–1136.

[110] Olof Arrhenius. "Species and area". In: *Journal of Ecology* 9.1 (1921), pp. 95–99.

[111] Mark Williamson, Kevin J Gaston, and WM Lonsdale. "The species–area relationship does not have an asymptote!" In: *Journal of Biogeography* 28.7 (2001), pp. 827–830.

[112] Mark V Lomolino. "Ecology's most general, yet protean pattern: The species-area relationship". In: *Journal of Biogeography* 27.1 (2000), pp. 17–26.

[113] John Harte and Justin Kitzes. "The use and misuse of species-area relationships in predicting climate-driven extinction". In: *Saving a Million Species*. Springer, 2012, pp. 73–86.

[114] John Harte, Adam B Smith, and David Storch. "Biodiversity scales from plots to biomes with a universal species–area curve". In: *Ecology letters* 12.8 (2009), pp. 789–797.

[115] Fangliang He and Stephen P Hubbell. "Species-area relationships always overestimate extinction rates from habitat loss". In: *Nature* 473.7347 (2011), pp. 368–371.

[116] Ann P Kinzig and John Harte. "Implications of endemics–area relationships for estimates of species extinctions". In: *Ecology* 81.12 (2000), pp. 3305–3311.

[117] David Storch, Petr Keil, and Walter Jetz. "Universal species-area and endemics-area relationships at continental scales". In: *Nature* 488.7409 (2012), pp. 78–81.

[118] Justin Kitzes and John Harte. "Beyond the species–area relationship: improving macroecological extinction estimates". In: *Methods in Ecology and Evolution* 5.1 (2014), pp. 1–8.

[119] Joel Rybicki and Ilkka Hanski. "Species–area relationships and extinctions caused by habitat loss and fragmentation". In: *Ecology letters* 16.s1 (2013), pp. 27–38.

[120] Claire Régnier et al. "Mass extinction in poorly known taxa". In: *Proceedings of the National Academy of Sciences* 112.25 (2015), pp. 7761–7766.

[121] Jared M Diamond, NP Ashmole, and PE Purves. "The present, past and future of human-caused extinctions [and discussion]". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 325.1228 (1989), pp. 469–477.

[122] Lian Pin Koh et al. "Species coextinctions and the biodiversity crisis". In: *science* 305.5690 (2004), pp. 1632–1634.

[123] Colin J Carlson et al. "Parasite biodiversity faces extinction and redistribution in a changing climate". In: *Science Advances* 3.9 (2017), e1602422.

[124] Robert R Dunn et al. "The sixth mass coextinction: are most endangered species parasites and mutualists?" In: *Proceedings of the Royal Society of London B: Biological Sciences* 276.1670 (2009), pp. 3037–3045.

[125] Giovanni Strona, Paolo Galli, and Simone Fattorini. "Fish parasites resolve the paradox of missing coextinctions". In: *Nature communications* 4 (2013), p. 1718.

[126] Daniel R Brooks and Eric P Hoberg. "How will global climate change affect parasite–host assemblages?" In: *Trends in parasitology* 23.12 (2007), pp. 571–574.

[127] Sabrina BL Araujo et al. "Understanding host-switching by ecological fitting". In: *PloS one* 10.10 (2015), e0139225.

[128] Eric R Dougherty et al. "Paradigms for parasite conservation". In: *Conservation Biology* (2015).

[129] Aaron M Ellison. "It's time to get real about conservation." In: *Nature* 538.7624 (2016), p. 141.

[130] Marten Scheffer et al. "Early-warning signals for critical transitions". In: *Nature* 461.7260 (2009), pp. 53–59.

[131] Sonia Kéfi et al. "Early warning signals of ecological transitions: methods for spatial patterns". In: *PloS one* 9.3 (2014), e92097.

[132] John M Drake and Blaine D Griffen. "Early warning signals of extinction in deteriorating environments". In: *Nature* 467.7314 (2010), p. 456.

[133] Christopher F Clements and Arpat Ozgul. "Including trait-based early warning signals helps predict population collapse". In: *Nature communications* 7 (2016).

[134] Carl Boettiger, Noam Ross, and Alan Hastings. "Early warning signals: the charted and uncharted territories". In: *Theoretical ecology* 6.3 (2013), pp. 255–264.

[135] Alan Hastings and Derin B Wysham. "Regime shifts in ecological systems can occur with no warning". In: *Ecology letters* 13.4 (2010), pp. 464–472.

[136] Christopher F Clements et al. "Body size shifts and early warning signals precede the historic collapse of whale stocks." In: *Nature ecology & evolution* 1.7 (2017), p. 188.

[137] Christopher F Clements et al. "Factors influencing the detectability of early warning signals of population collapse". In: *The American Naturalist* 186.1 (2015), pp. 50–58.

[138] Vasilis Dakos and Jordi Bascompte. "Critical slowing down as early warning for the onset of collapse in mutualistic communities". In: *Proceedings of the National Academy of Sciences* 111.49 (2014), pp. 17546–17551.

[139] Mark C Urban. "Accelerating extinction risk from climate change". In: *Science* 348.6234 (2015), pp. 571–573.

[140] John J Wiens. "Climate-related local extinctions are already widespread among plant and animal species". In: *PLOS Biology* 14.12 (2016), e2001104.

[141] Camille Parmesan and Gary Yohe. "A globally coherent fingerprint of climate change impacts across natural systems". In: *Nature* 421.6918 (2003), pp. 37–42.

[142] Gian-Reto Walther et al. "Ecological responses to recent climate change". In: *Nature* 416.6879 (2002), pp. 389–395.

[143] I-Ching Chen et al. "Rapid range shifts of species associated with high levels of climate warming". In: *Science* 333.6045 (2011), pp. 1024–1026.

[144] Aldina Franco et al. "Impacts of climate warming and habitat loss on extinctions at species' low-latitude range boundaries". In: *Global Change Biology* 12.8 (2006), pp. 1545–1553.

[145] Stavros D Veresoglou, John M Halley, and Matthias C Rillig. "Extinction risk of soil biota". In: *Nature communications* 6 (2015).

[146] Robert K Colwell, Robert R Dunn, and Nyeema C Harris. "Coextinction and persistence of dependent species in a changing world". In: *Annual Review of Ecology, Evolution, and Systematics* 43 (2012), pp. 183–203.

[147] Carlos Frederico Duarte Rocha, Helena Godoy Bergallo, and Emerson Brum Bittencourt. "More than just invisible inhabitants: Parasites are important but neglected components of the biodiversity". In: *Zoologia (Curitiba)* 33.3 (2016).

[148] Robert Poulin et al. "Missing links: testing the completeness of host-parasite checklists". In: *Parasitology* 143.1 (2016), pp. 114–122.

[149] JM Olwoch et al. "Climate change and the genus Rhipicephalus (Acari: Ixodidae) in Africa". In: *Onderstepoort Journal of Veterinary Research* 74.1 (2007), pp. 45–72.

[150] Graeme S Cumming and Detlef P Van Vuuren. "Will climate change affect ectoparasite species ranges?" In: *Global Ecology and Biogeography* 15.5 (2006), pp. 486–497.

[151] Sonia Altizer et al. "Climate change and infectious diseases: from evidence to a predictive framework". In: *science* 341.6145 (2013), pp. 514–519.

[152] Jonathan A Patz et al. "Global climate change and emerging infectious diseases". In: *Jama* 275.3 (1996), pp. 217–223.

[153] Kevin D Lafferty. "The ecology of climate change and infectious diseases". In: *Ecology* 90.4 (2009), pp. 888–900.

[154] Emily M York, Christopher J Butler, and Wayne D Lord. "Global decline in suitable habitat for Angiostrongylus (= Parastrongylus) cantonensis: the role of climate change". In: *PloS one* 9.8 (2014), e103831.

[155] Nyeema C Harris and Robert R Dunn. "Species loss on spatial patterns and composition of zoonotic parasites". In: *Proceedings of the Royal Society of London B: Biological Sciences* 280.1771 (2013), p. 20131847.

[156] Joseph A Cook et al. "Transformational principles for NEON sampling of mammalian parasites and pathogens: A response to Springer and colleagues". In: *BioScience* 66.11 (2016), pp. 917–919.

[157] GS Cumming. "Host preference in African ticks (Acari: Ixodida): a quantitative data set". In: *Bulletin of Entomological Research* 88.4 (1998), pp. 379–406.

[158] Graeme S Cumming. "Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida)". In: *Journal of biogeography* 27.2 (2000), pp. 425–440.

[159] Jorge Doña et al. "Global associations between birds and vane-dwelling feather mites". In: *Ecology* 97.11 (2016), pp. 3242–3242.

[160] DI Gibson, RA Bray, and EA Harris. *Host-parasite database of the natural history museum, London*. 2005.

[161] Giovanni Strona and Kevin D Lafferty. "FishPEST: an innovative software suite for fish parasitologists". In: *Trends in parasitology* 28.4 (2012), p. 123.

[162] JN Caira, K Jensen, and E Barbeau. "Global cestode database". In: *World Wide Web electronic publication. http://tapewormdb. uconn. edu* (2012).

[163] Arthur D Chapman and John Wieczorek. "Guide to best practices for georeferencing". In: *Copenhagen: Global Biodiversity Information Facility* (2006), pp. 1–77.

[164] NE Rios and HL Bart. "GEOLocate (Version 3.22)[computer software]". In: *Belle Chasse, LA: Tulane University Museum of Natural History* (2010).

[165] David Sánchez-Fernández, Jorge M Lobo, and Olga Lucía Hernández-Manrique. "Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles". In: *Diversity and Distributions* 17.1 (2011), pp. 163–171.

[166] David RB Stockwell and A Townsend Peterson. "Effects of sample size on accuracy of species distribution models". In: *Ecological modelling* 148.1 (2002), pp. 1–13.

[167] Pilar A Hernandez et al. "The effect of sample size and species characteristics on performance of different species distribution modeling methods". In: *Ecography* 29.5 (2006), pp. 773–785.

[168] Mary Suzanne Wisz et al. "Effects of sample size on the performance of species distribution models". In: *Diversity and distributions* 14.5 (2008), pp. 763–773.

[169] André SJ Proosdij et al. "Minimum required number of specimen records to develop accurate species distribution models". In: *Ecography* 39.6 (2016), pp. 542–552.

[170] Catherine H Graham et al. "The influence of spatial errors in species occurrence data used in distribution models". In: *Journal of Applied Ecology* 45.1 (2008), pp. 239–247.

[171] RJ Hijmans et al. *The WorldClim interpolated global terrestrial climate surfaces. Version 1.3*. 2004.

[172] Janet Franklin. *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, 2010.

[173] A Townsend Peterson. *Ecological niches and geographic distributions (MPB-49)*. 49. Princeton University Press, 2011.

[174] Greg J McInerny and Rampal S Etienne. "Ditch the niche–is the niche a useful concept in ecology or species distribution modelling?" In: *Journal of Biogeography* 39.12 (2012), pp. 2096–2102.

[175] Greg J McInerny and Rampal S Etienne. "Pitch the niche–taking responsibility for the concepts we use in ecology and species distribution modelling". In: *Journal of Biogeography* 39.12 (2012), pp. 2112–2118.

[176] Greg J McInerny and Rampal S Etienne. "Stitch the niche–a practical philosophy and visual schematic for the niche concept". In: *Journal of Biogeography* 39.12 (2012), pp. 2103–2111.

[177] Andrew J Davis et al. "Making mistakes when predicting shifts in species range in response to global warming". In: *Nature* 391.6669 (1998), pp. 783–786.

[178] Richard G Pearson and Terence P Dawson. "Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful?" In: *Global ecology and biogeography* 12.5 (2003), pp. 361–371.

[179] Miguel B Araujo et al. "Validation of species–climate impact models under climate change". In: *Global Change Biology* 11.9 (2005), pp. 1504–1513.

[180] JR Malenke, N Newbold, and DH Clayton. "Condition-specific competition governs the geographic distribution and diversity of ectoparasites". In: *The American Naturalist* 177.4 (2011), pp. 522–534.

[181] Damian K Dowling et al. "Feather mite loads influenced by salt exposure, age and reproductive stage in the Seychelles warbler Acrocephalus sechellensis". In: *Journal of Avian Biology* 32.4 (2001), pp. 364–369.

[182] V Dubunin. "Feather mites (Analgesoidea). Part I. Introduction to their study". In: *Fauna USSR* 6 (1951), pp. 1–363.

[183] Leandro Meléndez et al. "Climate-driven variation in the intensity of a host-symbiont animal interaction along a broad elevation gradient". In: *PloS one* 9.7 (2014), e101942.

[184] Shelby J Hiestand, Clayton K Nielsen, F Agustín Jiménez, et al. "Modelling potential presence of metazoan endoparasites of bobcats (Lynx rufus) using verified records". In: *Folia Parasitol* 61 (2014), pp. 401–410.

[185] GS Cumming. "Host distributions do not limit the species ranges of most African ticks (Acari: Ixodida)". In: *Bulletin of Entomological Research* 89.4 (1999), pp. 303–327.

[186] Sean P Maher et al. "Range-wide determinants of plague distribution in North America". In: *The American journal of tropical medicine and hygiene* 83.4 (2010), pp. 736–742.

[187] Andrés Lira-Noriega and A Townsend Peterson. "Range-wide ecological niche comparisons of parasite, hosts and dispersers in a vector-borne plant parasite system". In: *Journal of biogeography* 41.9 (2014), pp. 1664–1673.

[188] Wilfried Thuiller et al. "Biodiversity conservation: uncertainty in predictions of extinction risk". In: *Nature* 430.6995 (2004).

[189] Steven J Phillips, Miroslav Dudík, and Robert E Schapire. "A maximum entropy approach to species distribution modeling". In: *Proceedings of the twenty-first international conference on Machine learning*. ACM. 2004, p. 83.

[190] Steven J Phillips, Robert P Anderson, and Robert E Schapire. "Maximum entropy modeling of species geographic distributions". In: *Ecological modelling* 190.3 (2006), pp. 231–259.

[191] Jane Elith et al. "Novel Methods Improve Prediction of Species' Distributions from Occurrence Data". In: *Ecography* (2006), pp. 129–151.

[192] Charles B Yackulic et al. "Presence-only modelling using MAXENT: when can we trust the inferences?" In: *Methods in Ecology and Evolution* 4.3 (2013), pp. 236–243.

[193] Heather M Kharouba, Adam C Algar, and Jeremy T Kerr. "Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment". In: *Ecology* 90.8 (2009), pp. 2213–2222.

[194] Jane Elith, Michael Kearney, and Steven Phillips. "The art of modelling range-shifting species". In: *Methods in ecology and evolution* 1.4 (2010), pp. 330–342.

[195] Robert J Hijmans and Catherine H Graham. "The ability of climate envelope models to predict the effect of climate change on species distributions". In: *Global change biology* 12.12 (2006), pp. 2272–2281.

[196] Cory Merow, Matthew J Smith, and John A Silander. "A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter". In: *Ecography* 36.10 (2013), pp. 1058–1069.

[197] Mindy M Syfert, Matthew J Smith, and David A Coomes. "The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models". In: *PloS one* 8.2 (2013), e55158.

[198] Dan L Warren and Stephanie N Seifert. "Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria". In: *Ecological Applications* 21.2 (2011), pp. 335–342.

[199] Aleksandar Radosavljevic and Robert P Anderson. "Making better Maxent models of species distributions: complexity, overfitting and evaluation". In: *Journal of biogeography* 41.4 (2014), pp. 629–643.

[200] Robert Muscarella et al. "ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models". In: *Methods in Ecology and Evolution* 5.11 (2014), pp. 1198–1205.

[201] Kenneth P Burnham and David R Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[202] Mariya Shcheglovitova and Robert P Anderson. "Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes". In: *Ecological Modelling* 269 (2013), pp. 9–17.

[203] Jorge M Lobo, Alberto Jiménez-Valverde, and Raimundo Real. "AUC: a misleading measure of the performance of predictive distribution models". In: *Global ecology and Biogeography* 17.2 (2008), pp. 145–151.

[204] A Townsend Peterson, Monica Papeş, and Jorge Soberón. "Rethinking receiver operating characteristic analysis applications in ecological niche modeling". In: *Ecological modelling* 213.1 (2008), pp. 63–72.

[205] Samuel D Veloz. "Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models". In: *Journal of Biogeography* 36.12 (2009), pp. 2290–2299.

[206] Alissar Cheaib et al. "Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty". In: *Ecology letters* 15.6 (2012), pp. 533–544.

[207] Jason L Brown and Anne D Yoder. "Shifting ranges and conservation challenges for lemurs in the face of climate change". In: *Ecology and evolution* 5.6 (2015), pp. 1131–1142.

[208] Bernard WT Coetzee et al. "Ensemble models predict Important Bird Areas in southern Africa will become less effective for conserving endemic birds under climate change". In: *Global Ecology and Biogeography* 18.6 (2009), pp. 701–710.

[209] Laure Gallien et al. "Invasive species distribution models–how violating the equilibrium assumption can create new insights". In: *Global Ecology and Biogeography* 21.11 (2012), pp. 1126–1136.

[210] Ben L Phillips et al. "Parasites and pathogens lag behind their host during periods of host range advance". In: *Ecology* 91.3 (2010), pp. 872–881.

[211] Julie V Hopper et al. "Reduced parasite diversity and abundance in a marine whelk in its expanded geographical range". In: *Journal of biogeography* 41.9 (2014), pp. 1674–1684.

[212] Rob SA Pickles et al. "Predicting shifts in parasite distribution with climate change: a multitrophic level approach". In: *Global change biology* 19.9 (2013), pp. 2645–2654.

[213] John Harte et al. “Biodiversity conservation: Climate change and extinction risk”. In: *Nature* 430.6995 (2004).

[214] Samuel M Jantz et al. “Future habitat loss and extinctions driven by land-use change in biodiversity hotspots under four scenarios of climate-change mitigation”. In: *Conservation Biology* 29.4 (2015), pp. 1122–1131.

[215] A Townsend Peterson et al. “Future projections for Mexican faunas under global climate change scenarios”. In: *Nature* 416.6881 (2002), pp. 626–629.

[216] Jian Zhang et al. “Extinction risk of North American seed plants elevated by climate and land-use change”. In: *Journal of Applied Ecology* 54.1 (2017), pp. 303–312.

[217] Mariah E Hopkins and Charles L Nunn. “A global gap analysis of infectious agents in wild primates”. In: *Diversity and Distributions* 13.5 (2007), pp. 561–572.

[218] Sonja Matthee et al. “Epifaunistic arthropod parasites of the four-striped mouse, Rhabdomys pumilio, in the Western Cape Province, South Africa”. In: *Journal of Parasitology* 93.1 (2007), pp. 47–59.

[219] Vernon E Thatcher et al. “Amazon fish parasites.” In: *Amazoniana* 11.3/4 (1991), pp. 263–572.

[220] JL Luque and R Poulin. “Metazoan parasite species richness in Neotropical fishes: hotspots and the geography of biodiversity”. In: *Parasitology* 134.6 (2007), pp. 865–878.

[221] Regan Early and Dov F Sax. “Climatic niche shifts between species’ native and naturalized ranges raise concern for ecological forecasts during invasions and climate change”. In: *Global Ecology and Biogeography* 23.12 (2014), pp. 1356–1365.

[222] Eric P Hoberg and Daniel R Brooks. “Evolution in action: climate change, biodiversity dynamics and emerging infectious disease”. In: *Phil. Trans. R. Soc. B* 370.1665 (2015), p. 20130553.

[223] Eric P Hoberg and Daniel R Brooks. “Beyond vicariance: integrating taxon pulses, ecological fitting, and oscillation in evolution and historical biogeography”. In: *The biogeography of host-parasite interactions* (2010), pp. 7–20.

[224] Norman Myers et al. “Biodiversity hotspots for conservation priorities”. In: *Nature* 403.6772 (2000), pp. 853–858.

[225] Susan J Kutz et al. “”Emerging” Parasitic Infections in Arctic Ungulates”. In: *Integrative and Comparative Biology* 44.2 (2004), pp. 109–118.

[226] Lydden Polley, Eric Hoberg, and Susan Kutz. “Climate change, parasites and shifting boundaries”. In: *Acta Veterinaria Scandinavica* 52.1 (2010), S1.

[227] Rebecca Davidson et al. “Arctic parasitology: why should we care?” In: *Trends in Parasitology* 27.6 (2011), pp. 239–245.

[228] Jean Dupouy-Camet. "Parasites of cold climates: A danger or in danger?" In: *Food and Waterborne Parasitology* 4 (2016), pp. 1–3.

[229] Michela Pacifici et al. "Assessing species vulnerability to climate change". In: *Nature Climate Change* 5.3 (2015), pp. 215–224.

[230] Jessica C Stanton et al. "Warning times for species extinctions due to climate change". In: *Global change biology* 21.3 (2015), pp. 1066–1077.

[231] Abigail E Cahill et al. "How does climate change cause extinction?" In: *Proc. R. Soc. B*. The Royal Society. 2012, rspb20121890.

[232] Lance A Durden and James E Keirans. "Host–parasite coextinction and the plight of tick conservation". In: *American Entomologist* 42.2 (1996), pp. 87–91.

[233] Andrés Gómez, Elizabeth S Nichols, and Susan L Perkins. "Parasite conservation, conservation medicine, and ecosystem health". In: *New directions in conservation medicine: Applied cases of ecological health* (2012), pp. 67–81.

[234] Ismael Galván et al. "Feather mites (Acari: Astigmata) and body condition of their avian hosts: a large correlative study". In: *Journal of Avian Biology* 43.3 (2012), pp. 273–279.

[235] Mark Blaxter and Georgios Koutsovoulos. "The evolution of parasitism in Nematoda". In: *Parasitology* 142.S1 (2015), S26–S39.

[236] Roger Jovani et al. "Opening the Doors of Parasitology Journals to Other Symbionts". In: *Trends in Parasitology* (2017).

[237] RD Gregory. "Parasites and host geographic range as illustrated by waterfowl". In: *Functional Ecology* (1990), pp. 645–654.

[238] Peter W Price and Karen M Clancy. "Patterns in number of helminth parasite species in freshwater fishes". In: *The Journal of Parasitology* (1983), pp. 449–454.

[239] Naomi Attar. "ZIKA virus circulates in new regions". In: *Nature Reviews Microbiology* 14.2 (2016), pp. 62–62.

[240] GWA Dick, SF Kitchen, and AJ Haddow. "Zika virus (I). Isolations and serological specificity". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 46.5 (1952), pp. 509–520.

[241] Morgan Hennessey, Marc Fischer, and J Erin Staples. "Zika virus spreads to new areas-region of the Americas, May 2015–January 2016". In: *American Journal of Transplantation* 16.3 (2016), pp. 1031–1034.

[242] D Musso, EJ Nilles, and V-M Cao-Lormeau. "Rapid spread of emerging Zika virus in the Pacific area". In: *Clinical Microbiology and Infection* 20.10 (2014).

[243] Didier Musso, Duane J Gubler, et al. "Zika virus: following the path of dengue and chikungunya?" In: *The Lancet* 386.9990 (2015), pp. 243–244.

[244] Derek Gatherer and Alain Kohl. "Zika virus: a previously slow pandemic spreads rapidly through the Americas". In: *Journal of General Virology* 97.2 (2016), pp. 269–273.

[245] Dyan J Summers, Rebecca Wolfe Acosta, and Alberto M Acosta. "Zika virus in an American recreational traveler". In: *Journal of travel medicine* 22.5 (2015), pp. 338–340.

[246] Kevin Fonseca et al. "First case of Zika virus infection in a returning Canadian traveler". In: *The American journal of tropical medicine and hygiene* 91.5 (2014), pp. 1035–1038.

[247] Jason C Kwong, Julian D Druce, and Karin Leder. "Zika virus infection acquired during brief travel to Indonesia". In: *The American journal of tropical medicine and hygiene* 89.3 (2013), pp. 516–517.

[248] L Zammarchi et al. "Zika virus infection in a traveller returning to Europe from Brazil, March 2015". In: *Euro Surveill* 20.23 (2015), p. 21153.

[249] Nuno Rodrigues Faria et al. "Zika virus in the Americas: early epidemiological and genetic findings". In: *Science* 352.6283 (2016), pp. 345–349.

[250] Shlomit Paz and Jan C Semenza. "El Niño and climate change–contributing factors in the dispersal of Zika virus in the Americas?" In: *The Lancet* 387.10020 (2016), p. 745.

[251] A Townsend Peterson. "Ecologic niche modeling and spatial patterns of disease transmission". In: *Emerging infectious diseases* 12.12 (2006), p. 1822.

[252] Matthew E Aiello-Lammens et al. "spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models". In: *Ecography* 38.5 (2015), pp. 541–545.

[253] Moritz UG Kraemer et al. "The global distribution of the arbovirus vectors Aedes aegypti and Ae. albopictus". In: *Elife* 4 (2015), e08347.

[254] Moritz UG Kraemer et al. "The global compendium of Aedes aegypti and Ae. albopictus occurrence". In: *Scientific Data* 2 (2015), sdata201535.

[255] Jane P Messina et al. "A global compendium of human dengue virus occurrence". In: *Scientific Data* 1 (2014), p. 140004.

[256] Samir Bhatt et al. "The global distribution and burden of dengue". In: *Nature* 496.7446 (2013), pp. 504–507.

[257] Andrew D Haddow et al. "Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage". In: *PLoS neglected tropical diseases* 6.2 (2012), e1477.

[258] Scott C Weaver et al. "Zika virus: History, emergence, biology, and prospects for control". In: *Antiviral research* 130 (2016), pp. 69–80.

[259] Robert J Hijmans et al. "Very high resolution interpolated climate surfaces for global land areas". In: *International journal of climatology* 25.15 (2005), pp. 1965–1978.

[260] A Townsend Peterson et al. "Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 99.9 (2005), pp. 647–655.

[261] *NASA Earth Observations TERRA/MODIS NDVI*. `http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD13A2_M_NDVI`. Accessed: 2016-02-07.

[262] Wilfried Thuiller et al. "BIOMOD–a platform for ensemble forecasting of species distributions". In: *Ecography* 32.3 (2009), pp. 369–373.

[263] Paul Shapshak et al. "Zika virus". In: *Global Virology I-Identifying and Investigating Viral Diseases*. Springer, 2015, pp. 477–500.

[264] Chris Jones et al. "Twenty-first-century compatible CO2 emissions and airborne fraction simulated by CMIP5 earth system models under four representative concentration pathways". In: *Journal of Climate* 26.13 (2013), pp. 4398–4413.

[265] O Broennimann et al. *Ecospat: spatial ecology: miscellaneous methods. R package version 1.0*. 2014.

[266] Olivier Broennimann et al. "Measuring ecological niche overlap from occurrence and spatial environmental data". In: *Global Ecology and Biogeography* 21.4 (2012), pp. 481–497.

[267] Olivier Broennimann and Antoine Guisan. "Predicting current and future biological invasions: both native and invaded ranges matter". In: *Biology Letters* 4.5 (2008), pp. 585–589.

[268] David M Pigott et al. "Mapping the zoonotic niche of Ebola virus disease in Africa". In: *Elife* 3 (2014), e04395.

[269] Robert S Lanciotti et al. "Phylogeny of Zika virus in western hemisphere, 2015". In: *Emerging infectious diseases* 22.5 (2016), p. 933.

[270] Isaac I Bogoch et al. "Anticipating the international spread of Zika virus from Brazil." In: *Lancet (London, England)* 387.10016 (2016), pp. 335–336.

[271] Rebecca C Christofferson. "Zika virus emergence and expansion: lessons learned from dengue and chikungunya may not provide all the answers". In: *The American journal of tropical medicine and hygiene* 95.1 (2016), pp. 15–18.

[272] Jane P Messina et al. "The many projected futures of dengue". In: *Nature Reviews Microbiology* 13.4 (2015), pp. 230–239.

[273] Lars Eisen and Chester G Moore. "Aedes (Stegomyia) aegypti in the continental United States: a vector at the cool margin of its geographic range". In: *Journal of medical entomology* 50.3 (2013), pp. 467–478.

[274] Oliver J Brady et al. "Refining the global spatial limits of dengue virus transmission by evidence-based consensus". In: *PLoS neglected tropical diseases* 6.8 (2012), e1760.

[275] Michael A Johansson, Derek AT Cummings, and Gregory E Glass. "Multiyear climate variability and dengue–El Nino southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis". In: *PLoS medicine* 6.11 (2009), e1000168.

[276] Scott A Ritchie and Wayne Rochester. "Wind-blown mosquitoes and introduction of Japanese encephalitis into Australia." In: *Emerging infectious diseases* 7.5 (2001), p. 900.

[277] Medhat A Darwish et al. "A sero-epidemiological survey for certain arboviruses (Togaviridae) in Pakistan". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 77.4 (1983), pp. 442–445.

[278] Sophie Ioos et al. "Current Zika virus epidemiology and recent epidemics". In: *Medecine et maladies infectieuses* 44.7 (2014), pp. 302–307.

[279] Gabriele Neumann, Takeshi Noda, and Yoshihiro Kawaoka. "Emergence and pandemic potential of swine-origin H1N1 influenza virus". In: *Nature* 459.7249 (2009), pp. 931–939.

[280] Brian D Foy et al. "Probable non–vector-borne transmission of Zika virus, Colorado, USA". In: *Emerging infectious diseases* 17.5 (2011), p. 880.

[281] Feng Gao et al. "Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes". In: *Nature* 397.6718 (1999), pp. 436–441.

[282] Jorge A Alfaro-Murillo et al. "A cost-effectiveness tool for informing policies on Zika virus control". In: *PLoS neglected tropical diseases* 10.5 (2016), e0004743.

[283] Stephen S Morse et al. "Prediction and prevention of the next pandemic zoonosis". In: *The Lancet* 380.9857 (2012), pp. 1956–1965.

[284] Theo H Jetten and Dana A Focks. "Potential changes in the distribution of dengue transmission under climate warming." In: *The American journal of tropical medicine and hygiene* 57.3 (1997), pp. 285–297.

[285] Luis E Escobar and Meggan E Craft. "Advances and limitations of disease biogeography using ecological niche modeling". In: *Frontiers in Microbiology* 7 (2016).

[286] Pedro Segurado and Miguel B Araujo. "An evaluation of methods for modelling species distributions". In: *Journal of Biogeography* 31.10 (2004), pp. 1555–1568.

[287] Mathieu Marmion et al. "Evaluation of consensus methods in predictive species distribution modelling". In: *Diversity and distributions* 15.1 (2009), pp. 59–69.

[288] Leah R Johnson et al. "Understanding uncertainty in temperature effects on vector-borne disease: a Bayesian approach". In: *Ecology* 96.1 (2015), pp. 203–213.

[289] Michael Kearney and Warren Porter. "Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges". In: *Ecology letters* 12.4 (2009), pp. 334–350.

[290] Erin A Mordecai et al. "Detecting the impact of temperature on transmission of Zika, dengue, and chikungunya using mechanistic models". In: *PLoS neglected tropical diseases* 11.4 (2017), e0005568.

[291] Michael R Kearney, Brendan A Wintle, and Warren P Porter. "Correlative and mechanistic models of species distribution provide congruent forecasts under climate change". In: *Conservation Letters* 3.3 (2010), pp. 203–213.

[292] Colin Carlson, Eric Dougherty, and Wayne Getz. "An ecological assessment of the pandemic threat of Zika virus". In: *PLoS Neglected Tropical Diseases* 10.8 (2016), e0004968.

[293] PAHO/WHO. *Zika Epidemiological Report: Brazil. March 2017.* Tech. rep. Washington, D.C.: Pan-American Health Organization, 2017.

[294] PAHO/WHO. *Zika suspected and confirmed cases reported by countries and territories in the Americas Cumulative cases, 2015-2017. Updated as of 6 April 2017.* Tech. rep. Washington, D.C.: Pan-American Health Organization, 2017.

[295] Torsten J Naucke and Susanne Lorentz. "First report of venereal and vertical transmission of canine leishmaniosis from naturally infected dogs in Germany". In: *Parasites & vectors* 5.1 (2012), p. 67.

[296] Abdallah M Samy et al. "Mapping the global geographic potential of Zika virus spread". In: *Mem Inst Oswaldo Cruz* 111 (9 2016), pp. 559–560.

[297] Bruce Y Lee et al. "The potential economic burden of Zika in the continental United States". In: *PLoS neglected tropical diseases* 11.4 (2017), e0005531.

[298] Lauren A Castro et al. "Assessing real-time Zika risk in the United States". In: *BMC Infectious Diseases* 17.1 (2017), p. 284.

[299] Carrie A Manore et al. "Defining the risk of Zika and chikungunya virus transmission in human population centers of the eastern United States". In: *PLoS neglected tropical diseases* 11.1 (2017), e0005255.

[300] Michelle V Evans et al. "Data-driven identification of potential Zika virus vectors". In: *eLife* 6 (2017).

[301] Natasha Lindstrom. "Zika cases top 200 in Pennsylvania". In: *Tribune Live* (). URL: https://shar.es/1QKP8O.

[302] Dan Rivoli Chelsia Rose Marcius. "Zika virus prompts MTA plans for mosquito-killing larvicide in subway's stagnant pools of water, Cuomo announces". In: *New York Daily News* (). URL: http://nydn.us/2axoHxO.

[303] Alan Blinder. "Aimed at Zika mosquitoes, spray kills millions of honeybees". In: *New York Times* (). URL: https://nyti.ms/2olguVh.

[304] Seth Blumberg and James O Lloyd-Smith. "Inference of R_0 and transmission heterogeneity from the size distribution of stuttering chains". In: *PLoS Comput Biol* 9.5 (2013), e1002993.

[305] Isaac I Bogoch et al. "Potential for Zika virus introduction and transmission in resource-limited countries in Africa and the Asia-Pacific region: a modelling study". In: *The Lancet Infectious Diseases* 16.11 (2016), pp. 1237–1245.

[306] T Alex Perkins et al. "Model-based projections of Zika virus infections in childbearing women in the Americas". In: *Nature microbiology* 1 (2016), p. 16126.

[307] Alberto J Alaniz, Antonella Bacigalupo, and Pedro E Cattan. "Spatial quantification of the world population potentially exposed to Zika virus". In: *International Journal of Epidemiology* (2017), dyw366.

[308] Justin Lessler et al. "Assessing the global threat from Zika virus". In: *Science* 353.6300 (2016), aaf8160.

[309] Enbal Shacham et al. "Potential High-Risk Areas for Zika Virus Transmission in the Contiguous United States". In: *American Journal of Public Health* (2017), e1–e8.

[310] José Santos and Bruno M Meneses. "An integrated approach for the assessment of the Aedes aegypti and Aedes albopictus global spatial distribution, and determination of the zones susceptible to the development of Zika virus". In: *Acta tropica* 168 (2017), pp. 80–90.

[311] Erin A Mordecai et al. "Optimal temperature for malaria transmission is dramatically lower than previously predicted". In: *Ecology letters* 16.1 (2013), pp. 22–30.

[312] Daozhou Gao et al. "Prevention and control of Zika as a mosquito-borne and sexually transmitted disease: a mathematical modeling analysis". In: *Scientific reports* 6 (2016).

[313] Francesca Pozzi, Christopher Small, and Gregory Yetman. "Modeling the distribution of human population with nighttime satellite imagery and gridded population of the world". In: *Earth Observation Magazine* 12.4 (2003), pp. 24–30.

[314] Richard Levins. "The strategy of model building in population biology". In: *American scientist* 54.4 (1966), pp. 421–431.

[315] Mark Dickey-Collas et al. "Hazard warning: model misuse ahead". In: *ICES Journal of Marine Science* 71.8 (2014), pp. 2300–2306.

[316] Felipe De Jesus Colón-González et al. "After the epidemic: Zika virus projections for Latin America and the Caribbean". In: *PLoS Neglected Tropical Diseases* (2017).

[317] Nathan D Grubaugh et al. "Genomic epidemiology reveals multiple introductions of Zika virus into the United States". In: *Nature* 546.7658 (2017), pp. 401–405.

[318] Grace HY Leung et al. "Zika virus infection in Australia following a monkey bite in Indonesia". In: *Southeast Asian Journal of Tropical Medicine and Public Health* 46.3 (2015), p. 460.

[319] AWR McCrae and BG Kirya. "Yellow fever and Zika virus epizootics and enzootics in Uganda". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 76.4 (1982), pp. 552–562.

[320] Ricardo Lourenço-de-Oliveira and Anna-Bella Failloux. "High risk for chikungunya virus to initiate an enzootic sylvatic cycle in the tropical Americas". In: *PLoS neglected tropical diseases* 11.6 (2017), e0005698.

[321] Izabela K Ragan et al. "Investigating the Potential Role of North American Animals as Hosts for Zika Virus". In: *Vector-Borne and Zoonotic Diseases* 17.3 (2017), pp. 161–164.

[322] Marco Ajelli et al. "Host outdoor exposure variability affects the transmission and spread of Zika virus: Insights for epidemic control". In: *PLoS neglected tropical diseases* 11.9 (2017), e0005851.

[323] T Alex Perkins et al. "Heterogeneity, mixing, and the spatial scales of mosquito-borne pathogen transmission". In: *PLoS Comput Biol* 9.12 (2013), e1003327.

[324] Ian T Kracalik et al. "Modeling the environmental suitability of anthrax in Ghana and estimating populations at risk: Implications for vaccination and control". In: *PLOS Neglected Tropical Diseases* 11.10 (2017), e0005885.

[325] David W Redding et al. "Environmental-mechanistic modelling of the impact of global change on human zoonotic disease emergence: a case study of Lassa fever". In: *Methods in Ecology and Evolution* 7.6 (2016), pp. 646–655.

[326] David M Pigott et al. "Local, national, and regional viral haemorrhagic fever pandemic potential in Africa: a multistage analysis". In: *The Lancet* (2017).

[327] Duygu Balcan et al. "Seasonal transmission potential and activity peaks of the new influenza A (H1N1): a Monte Carlo likelihood analysis based on human mobility". In: *BMC medicine* 7.1 (2009), p. 45.

[328] Nicholas H Ogden et al. "Risk of travel-related cases of Zika virus infection is predicted by transmission intensity in outbreak-affected countries". In: *Parasites & Vectors* 10.1 (2017), p. 41.

[329] Flavio Codeço Coelho et al. "Higher incidence of Zika in adult women than adult men in Rio de Janeiro suggests a significant contribution of sexual transmission from men to women". In: *International Journal of Infectious Diseases* 51 (2016), pp. 128–132.

[330] Sherry Towers et al. "Estimate of the reproduction number of the 2015 Zika virus outbreak in Barranquilla, Colombia, and estimation of the relative role of sexual transmission". In: *Epidemics* 17 (2016), pp. 50–55.

[331] Laith Yakob et al. "Low risk of a sexually-transmitted Zika virus outbreak". In: *The Lancet infectious diseases* 16.10 (2016), pp. 1100–1102.

[332] David Baca-Carrasco and Jorge X Velasco-Hernández. "Sex, Mosquitoes and Epidemics: An Evaluation of Zika Disease Dynamics". In: *Bulletin of Mathematical Biology* 78.11 (2016), pp. 2228–2242.

[333] Antoine Allard et al. "The risk of sustained sexual transmission of Zika is underestimated". In: *bioRxiv* (2016), p. 090324.

[334] KM Folkers, AL Caplan, and LH Igel. "Zika, sexual transmission and prudent public health policy". In: *Public Health* 148 (2017), pp. 66–68.

[335] David G Regan and David P Wilson. "Modelling sexually transmitted infections: less is usually more for informing public health policy". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 102.3 (2008), pp. 207–208.

[336] Nick Golding et al. "The zoon R package for reproducible and shareable species distribution modelling". In: *Methods in Ecology and Evolution* (2017).

[337] Nikolaos Zacharias et al. "First Neonatal Demise with Travel-Associated Zika Virus Infection in the United States of America". In: *American Journal of Perinatology Reports* 7.02 (2017), e68–e73.

[338] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: `https://www.R-project.org`.